Philosophy, Religion and Education

Vol. 51 AUGUST 1960 No 4.

The Journal of Educational Psychology

CONTENTS

A Comparison of the Attitudes and Interests of Ninth-Grade Students over Two Decades 175
MARY COVER JONES
Short-Forms of the Wechsler Intelligence Scale for Children in the Educable, Noninstitutionalized, Mentally Retarded 187
LEWIS SCHWARTZ AND EUGENE E. LEVITT
Grade Differences in Transfer as a Function of the Time Interval between Learning Tasks 191
MICHAEL GLADIS
Stability of Item Statistics in Two Institutions 195
ERNEST R. HILGARD AND EARL L. JANDRON
The Effects of Verbal Mediation on the Modification of Children's Attitudes 199
SAMUEL H. OSIPOW
Sociability and Academic Achievement in Various Types of Learning Situations 208
LESLIE R. BEACH
Variation of Whole-Part Methods of Learning 213
F. J. MCGUIGAN
The Effects of Visual Discrimination Pretraining on Learning to Read a Vocabulary List in Kindergarten Children 217
SIEGMAR MUEHL
College Press and Changes in Study Plans of Talented Students 222
DONALD L. THISTLETHWAITE
Faculty Perceptions of Student Success in Teacher Education 235
ROGER A. MYERS
Changes in Scores on the Northwestern Misconceptions Test in Six Elementary Psychology Courses 240
W. J. MCKEACHIE

AMERICAN PSYCHOLOGICAL ASSOCIATION

The Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

LEE J. CRONBACH
University of Illinois
FRANCIS J. DIVESTA
Syracuse University
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College For Teachers
HAROLD E. JONES
University of California, Berkeley
IRVING LORGE
Teachers College, Columbia University
DOROTHEA MCCARTHY
Fordham University
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
PERCIVAL M. SYMONDS
Teachers College, Columbia University
GEORGE G. THOMPSON
Ohio State University

The Journal of Educational Psychology publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-Class postage paid at Baltimore, Md.

Copyright, 1960 by the American Psychological Association, Inc.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 51 August 1960 Number 4

A COMPARISON OF THE ATTITUDES AND INTERESTS OF NINTH-GRADE STUDENTS OVER TWO DECADES

MARY COVER JONES

Department of Education and Institute of Human Development, University of California, Berkeley

This paper compares the interests, activities and opinions of a sample of ninth graders in 1935 with those of ninth graders in the same junior high school in 1953[1] and again in 1959. The 1935 sample consists of members of the Oakland Growth Study (1932–1939). As adults, this group is now cooperating in a follow-up study in which records are being obtained of their contemporary interests and attitudes and their estimate of how they believe they answered similar items when they were adolescents. To evaluate the consistency of the records, it seems important to consider trends which may have occurred as a result of social changes. Tests of new groups of the same age as the adolescents in our original sample have provided the basis for this comparison survey.

Are teenagers today more serious or more frivolous, more conforming or more rebelling, more sophisticated or more naive, than teenagers of a generation ago? This is not a question which can have a general answer, since trends (if they occur) will not necessarily be in the same direction in different regions or in different social classes. However, where relevant data are available for known samples it seems desirable to report them and consider their implications in relation to other research in this field. What adolescents think, do, want is of interest from many different theoretical and practical points of view.

[1] The comparison of 1935 with 1953 data is more extensive, including sampling descriptions on the basis of census tract analysis of residence, intelligence test scores and ratings of parents' occupation. The 1959 testing was undertaken as a supplementary checking device for our 1953 interpretations and to bring the study up to date.

Comparison of Samples

In a state growing as rapidly as California, the population composition of some neighborhoods has changed so greatly that temporal trends would be difficult to disentangle from differences in social groups. Fortunately for our purpose, the neighborhoods involved in this study have to a fairly high degree maintained the stability which was noted at the time this junior high school was chosen for the longitudinal study.

Table 1 compares the distribution of parents' occupations (Edwards, 1933) in 1935 and 1953, and Table 2 the census tract classifications for home addresses of our two samples, based on Tryon's cluster analysis of the 1940 and 1950 census tract data for this area. An increase has occurred in professional and managerial occupations, and the total for white collar occupations has increased from 59 to 65%. Similarly, a slight upward shift has occurred in the neighborhood classifications; here the principal change has resulted from the opening of a neighborhood of new homes, and the

TABLE 1

SOCIOECONOMIC CLASSIFICATION: PARENTS' OCCUPATION (EDWARDS 6-POINT SCALE)
(Expressed in percentages)

	1935	1953
1. Professional	07	17
2. Managerial	18	29
3. Semiprofessional	38	25
4. Skilled	29	25
5. Semiskilled	06	04
6. Unskilled	02	00
Median	3.12	2.59
SD	1.04	1.12

TABLE 2

CENSUS TRACT CLASSIFICATION[a]
(Expressed in percentages)

	FAS	1935	1953
1. Exclusive[b]	768	21	25.6
2. Good neighborhood	666	38	41
3. New good neighborhood	576	00.6	14
4. Neighborhood of modest homes	466	29	17

[a] Tryon classification (S is high):
F = Family life, i.e., detached owner occupied home
A = Assimilation, i.e., native white
S = Socioeconomic independence, i.e., managerial, professional males

[b] The four classifications listed account for the majority of the subjects in this study.

diminished proportion remaining in an older neighborhood.[2]

To some extent these changes reflect the general upward mobility in living standards, income and education which has occurred over this 18-year period. An additional contributing factor was a change in the elementary school sources of this junior high school population. Before 1940 it drew from five elementary schools in the northern part of Oakland. By 1953 a new elementary school (mean occupational rating 2.59) had been added, and one school with an occupational rating of 3.9 had been displaced, its pupils being sent to a different junior high school. This change is in the direction of a more homogeneous middle class group, slightly above the former average.

If these changes had resulted in a very marked alteration of the social structure of the junior high school district, we would expect to find significant changes in the intelligence distribution. This has not occurred as is shown in Table 3.

As a further control on the comparability of the two populations in socioeconomic status, a subsample was selected of 40 cases for each sex and each period under study (1935 and 1953), carefully matched on the basis of parents' occupation. The responses for the subsamples on a 139 item comparison were in general agreement with those of the larger sample. A selection of these items, Things to be—Occupations is shown in Table 4. Because of the advantages of dealing with a larger sample, it was decided to use the total ninth grade population as the basis of this report.

TABLE 3

INTELLIGENCE TEST (IQ) CLASSIFICATION
(Expressed in percentages)

	1935[a]	1953[b]
130–139	02	01
120–129	06	12
110–119	27	30
100–109	38	30
90–99	21	18
80–89	05	07
70–79	01	02
Mean	105.49	105.18
SD	10.75	12.28

[a] Terman Group Test, Form A and B.
[b] Kuhlmann-Anderson.

[2] Tryon has compared eight social areas in his study of 1940 and 1950 census tract data in the San Francisco Bay Area, and has noted that "... in the year 1940, people who lived in these social areas differed greatly from each other." The differences (between these areas) in 1940 and in 1950 remain almost exactly the same (Tryon, 1955, 1959). In a personal communication Tryon has affirmed that the Oakland section from which our sample was drawn has remained highly stable by his method of analysis.

Subjects and Procedures

Results are discussed below for an inventory consisting of some 250 items pertaining to activities, interests, and attitudes. Levels of significance (Walker, 1953) have been computed for the sexes separately in a comparison of earlier with later records. Sex differences within each temporal sample have been analyzed, as well as age trends over the seven-year period in the earlier, longitudinal sample.

Members of the Oakland Growth study, 72 boys and 70 girls, are represented in the 1935 testing, when they were in the ninth grade. For this group an interest record (Jones, 1940) was administered yearly for seven years, an activity schedule yearly from the 9th through the 12th grade, and an opinion ballot in the 8th, 9th, and 11th grades.

For the 1953 comparison, all ninth-grade students in the same junior high school were tested. The boys numbered 95 and the girls 78. The additional test program in 1959 included 134 boys and 123 girls. Nonwhite students, now numbering about 20, were eliminated to maintain comparability with the earlier samples.

The test used in 1953 and in 1959 included a representative selection of items from the three inventories used in the longitudinal study. It was administered in one class period of 50 min. Students were asked to check items organized on separate pages under such headings as Where I Went Last Week, and What I Read Last Week. They were asked to indicate whether they liked, disliked, or were indifferent to a number of items listed under titles such as Things to Own, or Things to Be: Occupations. Headings such as Things You Talk About were answered on a 5-point scale from Never to Very often.[a] Attitudes were sampled under the heading My Opinion, which instructed the *Ss* to report whether they approved, disapproved, or sometimes approved the behavior or ideas listed.

TABLE 4

Things to Be—Occupations
(A Comparison of Responses for 40 Selected Cases and for the Total Sample)

	Boys % Positive Response			
	1935		1953	
	N = 40	*N* = 72	*N* = 40	*N* = 95
Architect	45	49	37.5	35
Athletic Coach	67.5	61	40	42
Artist	25	31	05	09
Aviator	67.5	68	42.5	44
Chemist	37.5	36	25	29
College Teacher	27.5	25	05	07
Dentist	15	17	27.5	21
Doctor	32.5	31	35	39
Engineer	52.5	53	60	55
Farmer	42.5	37	45	28
Lawyer	35	41	25	33
Policeman	27.5	28	17.5	23
	Girls % Positive Response			
	1935		1953	
	N = 40	*N* = 70	*N* = 40	*N* = 78
Artist	35	48	22.5	22
Bookkeeper	35	36	20	19
Decorator	20	30	45	44
Doctor	30	22	25	22
Kindergarten Teacher	35	33	30	35
Librarian	22.5	23	20	18
Nurse	42.5	41	35	36
Secretary	60	62	70	60
Stenographer	47.5	59	40	32
Teacher	25	30	22.5	24
Telephone Operator	32.5	31	32.5	29
Waitress	20	19	22.5	20

The categories of items as presented were administrative rather than functional. Items belonging to a given functional classification (such as interest in sex-social activities) occurred under different phrasing in each of the nine parts of the test. Thus, preoccupation with mixed group activities and interest in the opposite sex might presumably be tapped by the *Ss* desire to own

[a] Some of the items were adaptations and extensions of material included in tests by Furfey (1928), Lehman and Witty (1927), Pressey (1933), and Symonds (1928).

new party clothes, to go to a good dance hall, to want to stay out late, or to talk about dates. Similarly, a student might indicate that he or she went to a dance or read a love story last week, and approved of "permanents."

Results

One fact is outstanding in our results. In the more recent tests, boys and girls in the ninth grade marked items in such a way as to indicate greater maturity and greater social sophistication. This is demonstrated in their responses to items which show an age trend in adolescence for the earlier study group. Some of these items, representing "juvenile" interests or immature behavior, are, in the longitudinal sample, rejected increasingly as boys and girls grow older. At the same time, other items which represent greater maturity of interests and attitudes become more frequently selected. For example, members of the adolescent study group, as they grew older, no longer wanted to "own a stamp collection," to read "*Boy's Life*," to go to a "secret clubhouse" or to be "a detective." On the other hand, as they grew older, they more often checked such items as this week "I went to a dance"; I approve of "thinking about how I look"; I talk about "having dates" (Table 5).

On such age-relative items, we now find that ninth graders in the '50s are more

TABLE 5
Social-Sex Interests

	Boys % Positive Response 1935	1953	1959	Girls % Positive Response 1935	1953	1959
	N = 72	*N* = 70	*N* = 134	*N* = 70	*N* = 78	*N* = 123
Activities						
I took care of my clothes[a]	9	21*	22	70	86	77*
I manicured my nails	0	7*	16	62	74*	63*
I went to a dance	25	27*	40	24	41	61
I read a love story	8	3*	5*	30	42	55
Interests						
Would like to own a phonograph and dance records	55	60*	72	50	93	92
Conversations						
I talk about having dates[b]	46	56	63	58	92	84
I talk about clothes, things to wear	42	57	66	89	94*	91*
I talk about girl friends	55	71	74			
I talk about boy friends				65	94	96
Attitudes						
I approve of using lipstick[c]	23	66	66	58	99	96
I approve of "permanents"[d]	29	49	42*	62	72*	57*
I approve of thinking about how you look	27	51	65	40	77	72
I approve of love scenes in movies	26	29*	37*	39	59	53

[a] Check Yes
[b] Never, Seldom, Sometimes, Often, Very often. Per cent equals sum of Sometimes, Often, Very often.
[c] Yes, No, Sometimes.
[d] The reduced positive response for both sexes in 1959 probably reflects a change in hairdress fashion.
* Comparisons *not* significant at the .05 level or better 1935–1953 and 1935–1959 are starred. Significance levels were not determined for the positive response percentages shown in the table, but from chi squares based on *all* responses (e.g., Yes, No, Sometimes).

comparable to the eleventh- or twelfth-grade students of 20 years earlier than to those of their own age (Table 5).

Interest, activity, and attitude items which illustrate this tendency to check the more mature response have been grouped under subheadings such as social-sex interests, sense of responsibility, tolerance, and antiaggression attitudes. For each of these categories, one item which illustrates the findings has been selected for discussion.

In the area of social-sex or heterosexual preoccupation, Fig. 1 shows the response for the item *I approve of the use of lipstick*. For boys and girls respectively, the solid and broken lines indicate age trends in the longitudinal sample, in the 1930s. The responses of ninth graders in 1953 and 1959 are also indicated; black circles represent the percentages for boys, open circles the percentages for girls. Lipstick symbolizes as well as any one specific item could, the sensitization in early adolescence toward a new sex role and toward being grown up. While its use is confined to girls, opinions about its use are not, as the figure indicates. The lag shown in the data for boys (as compared with girls) may indicate either less interest or active antagonism toward signs of maturing from members of the less mature sex who are being pressured into the heterosexual phase ahead of schedule. Recorded conversations in the 1930s suggest the latter:

Boy to Girl: Where's Jane?

Girl: Down at the beauty parlor getting ready for the dance.

Boy: They all look alike to me whether they get fixed up or not.

Or going down the hall the first day of school, Roy says to Bob: There goes Irene. And the reply is: All painted up and looks like a horse.

Adult observers of seventh graders in the Growth Study considered it worth a notation that Shirley had lipstick on. In the high eighth grade, however, they commented: The Foster sisters were the only ones at the party without lipstick.

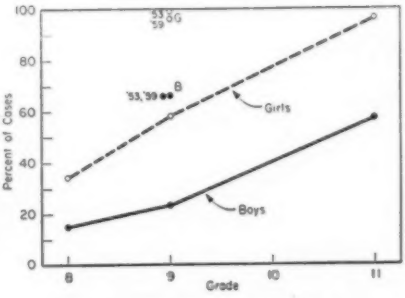

FIG. 1. Approve of using lipstick

The data show that by the 1950s (1953 and 1959), for both boys and girls, the percentage approving of the use of lipstick was higher in the 9th grade than in the 11th grade sample 20 years earlier. It is well known that there are social class and regional differences in attitudes toward the use of lipstick and other aspects of adolescent grooming. But here we find also a temporal difference in samples living in the same neighborhood and comparable in social status.

In a recent study comparing Minnesota junior and senior high school student's responses in 1957 to those of Symond's (1936) Tulsa and New York City youth in 1935, Harris (1959) interprets similar findings: "Today youth marry younger and show earlier interest in social relations, love and marriage." (p. 458) Our own evidence of interest in grooming activities among ninth graders shows how a temporal trend in heterosexual preoccupation may extend down into the middle period of adolescence.

Other items in the category of social-sex interest are shown in Table 5. For the most part (except for the few comparisons indicated by asterisks) differences are significant between 1935 and 1953 and between 1935 and 1959.

When we described the junior high school culture of our group in the 1930s, we interpreted their concern with social activities as in conflict with achievement motives of an academic nature. This preoccu-

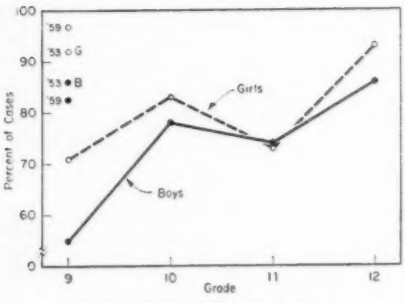

FIG. 2. This week I studied

pation often seemed inconsistent with an interest in homework, in assuming responsibilities or in showing concern for any but their own narrow peer group activities (Stolz, 1937). Does a similar situation hold today? Is an increased social interest accompanied by a decreased commitment in other areas?

Quite the contrary. These 14 year olds in 1953 and 1959 can be described from their answers as more studious, more broadly interested in the contemporary scene, more tolerant in their social attitudes, more inclined to value controlled behavior and to disapprove of aggression or irresponsible behavior in others. Comparative results for the item *This week I studied* are shown in Fig. 2. A significantly higher percentage of boys and of girls in the 1950s reported studying. The actual percentages in the ninth grade were virtually the same as those of high school seniors in the 1930s. Other items indicating a greater sense of personal and social responsibility, as marked in the more recent sample, are shown in Table 6. There were more frequent reports of talking about their studies and talking about other school activities. They liked to go to school assemblies. Such responses indicate a more school-centered orientation in the '50s than in the '30s.

A greater proportion said they earned money, took care of their clothes and their rooms, talked about what they were going to be when they grew up and thought it appropriate to worry about the future (Table 6). This, in prosperous times as contrasted with depression years and (for the 1953 sample) before Sputnik!

The differences shown in Table 6 may be in part differences in the *Zeitgeist*, and in part a matter of situational changes. Thus, an increase in the number of young children, and the development of baby-sitting, may be largely responsible for the increase (especially among girls) in "taking care of children" and "earning money." The increased interest reported by both sexes in such items as "reading about science" and "talking about political matters" may define a generally higher level of intellectual-cultural maturity in the more recent junior high school samples. The markedly significant increases especially among girls in religious interests may reflect an additional factor; the percentages reached in the '50s are higher than any we obtained even in the 12th grade in the '30s (Fig. 3).

Table 6 presents data for a number of items illustrating different aspects of tolerance. As shown in Figure 4, *I approve of foreigners* elicits a much more affirmative response in the '50s, from both boys and girls.

One group of items in Table 7 (approval of betting, gambling, etc.) suggests a more relaxed attitude which might be thought to be incompatible with the greater expressed interest in the church and its influences (as shown in the preceding table). However, this acquiescence does not extend to what might be classified as inconsiderate or aggressive behavior. The ninth graders of the '50s manifest more disapproval of antisocial behavior such as "*getting even*" or "doing what I want when I want to" (Fig. 5).

Some of the responses discussed above suggest a greater socialization and also a more conforming attitude. The later characteristic is frequently attributed to our present generation of young people. But not all items which connote acquiescence are approved by the recent samples as

TABLE 6

RESPONSIBLE BEHAVIOR

	Boys % Positive Response			Girls % Positive Response		
	1935	1953	1959	1935	1953	1959
Activities						
I studied	67	86	85	54	92	94
I took care of children	8	20	25	30	47	53
I earned some money	57	69*	69*	22	58	49
I read						
something about science	17	35	40	4	12*	17
The Bible	16	20*	21*	12	35	32
something about religion	8	19*	18*	11	28	30
I went to Sunday School	41	43*	40*	35	69	62
I went to Church	46	47*	56*	55	82	78
Conversations						
I talk about						
what you are going to be	58	77	69*	74	89	89
money and things you need	60	78	75	53	84	86
church and things about religion	21	46	45	33	72	71
studies, classwork	63	83	80	82	93	91
school activities	49	59*	64*	58	89	82
things about the government, politics	38	67	48*	14	51	30
Attitudes						
I Approve of						
worrying about the future	12	27	31	5	17	32

* Comparisons with 1935 are starred for chi squares *not* significant at the .05 level or better.

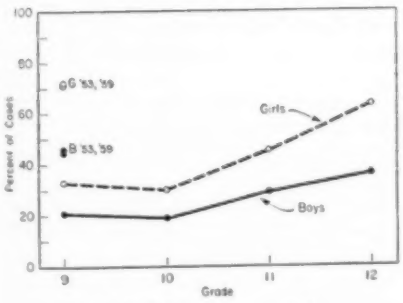

FIG. 3. Talking about: religion

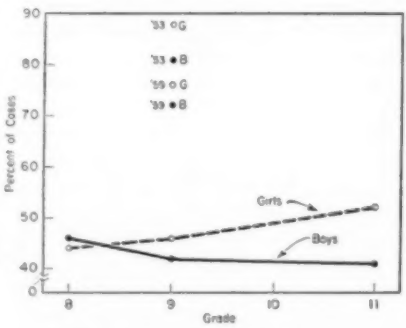

FIG. 4. Approve of foreigners

compared with the '30s. More adolescents countenance "disagreeing with my parents," and "saying exactly what I think." "Always doing what is expected," is indicated as much less acceptable.

More in line with the findings of the present study is an early report by Pressey (1946) on attitude changes between 1923 and 1943. Pressey was encouraged by "the general freeing of young people over the past 20 years . . . from a great variety of borderland social taboos, inhibitions and restrictions . . ." (p. 186). A second trend

TABLE 7
TOLERANT ATTITUDES

	Boys % Positive Response			Girls % Positive Response		
	1935	1953	1959	1935	1953	1959
Approving of foreigners	42	81	72	46	88	76
Giving criminals another chance	9	34	40	6	22	39
Approving of women doctors	44	72	61	76	90	88*
Tolerance toward						
gambling	6	21	34	2	7*	5
betting on games	25	40	47	17	13*	23*
playing cards	64	86	77	66	72*	71*
playing cards for money	18	33	34	11	5*	12*
women smoking	18	35	31	26	35*	22

* Comparisons with 1935 are starred for chi squares not significant at the .05 level or better.

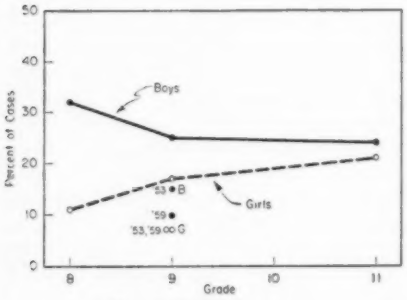

FIG. 5. Getting even

which he applauded was "the growth of activities" (p. 187).

DISCUSSION

There is little evidence in the recent literature to support the rather favorable contemporary picture of adolescents as reported here. Harris' 1957 study indicates that Minnesota young people are somewhat more concerned about study-habits than were students in 1935 but that they are not much "interested." Similarly, they seemed unconcerned about civic affairs. Remmers and Radler (1957) lament that teenagers, rebellious of adult authority in many areas, nevertheless tend to follow the example of adults in being illiberal and conformist.

Other studies based on attitudes questionnaires have found (among college students) considerable evidence of social conscience and of toleration for nonconforming behavior and ideas (Jacob, 1957; Sanford, 1956; Webster, 1958). At the same time these *Ss* were reported to show little concern about social issues.

The data summarized here are for a small sample in a single geographic location representing a limited range of social status. But if we believe what our *Ss* tell us, all is not lost for youngsters growing up in the '50s.

SEX DIFFERENCES AND SEX ROLE PATTERNS

Check lists of children's interests, such as those employed in this study, have been used by a number of investigators to examine the development of sex role concepts and sex differences. For most of the items in the area of social-sex development (listed in Table 5) the indication of mature interest is more marked in girls than in boys. More girls than boys report "going to a dance." More girls than boys report that they talk about dates and opposite-sex friends.

Earlier physiological maturing of girls (as compared with boys) would be expected to produce now, as formerly, an earlier maturing in social-sex development. However, it may be that the temporal trend in the direction of a generally earlier expression of heterosexual interest for both sexes, contributes to some blurring of sex role patterns as revealed in certain aspects of the data for the 1950s.

For example, this may account for the fact that in the 1950s boys' responses are more like those of girls on some items.

More girls than boys checked that they manicured their nails, but between 1935 and 1959 the increase in percentage checking this item was significant only for boys.[4]

The boys' reduced interest in some aspects of scientific and mechanical manipulation (usually considered a masculine interest)[5] may be a temporary expression of earlier social maturing. To indicate interest in experimenting with a chemistry set or with electricity may at a certain age seem juvenile rather than socially mature. The percentages for the item *experimenting with chemistry* are shown in Fig. 6. The drop in interest for ninth-grade boys is significant at the .05 level in 1953 and in the same direction, though not significant at that level, in 1959. The data from both testings in the '50s indicate less interest in such matters than at any age in the '30s, even though the longitudinal sample maintained a downward trend with age in items of this nature.

Although greater social maturity in both boys and girls may be one factor which accounts for increased similarity of the sexes on some items, there is one area of interest in which results run counter to this observation. The girls seem to have made a striking shift from passive to active interest in sports and athletic activities.[6] In the earlier sample the percentage checking the positive response diminished from the 6th to the 12th grade. There was also a large sex difference in the 1935 sample, with boys showing more favorable attitudes.

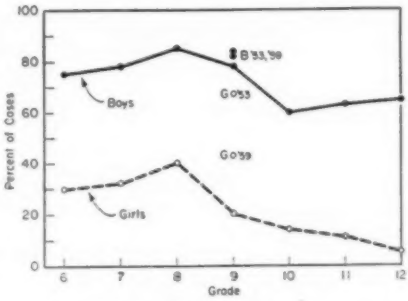

FIG. 6. Experiment with chemistry

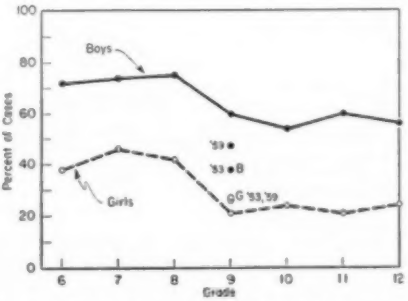

FIG. 7. Own a baseball and bat

Figure 7 gives the results for a representative item, showing the trend in percentage wishing *to own a baseball and bat*. Data are given for the 6th through the 12th grades in the '30s and also for the 9th grade in 1953 and 1959. Differences are significant for girls but not for boys (Table 8).

Recent studies by Rosenberg and Sutton-

[4] The writer, who administered the tests on both occasions, can corroborate from observational notes the evidence of increased grooming activities in boys. Increased hair grooming was especially in evidence.

[5] In 1958 Marshall used some of the items of the Interest Inventory to indicate awareness of appropriate or inappropriate sex-roles in fifth-grade boys and girls. Both boys and girls rated "experimenting with electricity" as something which boys would like and girls would dislike.

[6] This is the only finding in the present investigation which is not supported by the school personnel from their observation and experience. The same teacher has been in charge of the girls' physical education program since the early thirties; she could offer no explanation of any results on the basis of curriculum change or emphasis. However, several teachers and administrators cited the effective city recreational program which has been built up over this period, as a possible contributing factor.

A study by Meleny (1959) comparing the motor performance of eighth-grade girls now as compared to our longitudinal sample 24 years ago found no striking differences in these abilities. The more recent sample however was significantly taller and heavier.

TABLE 8
INTEREST IN ATHLETICS

	Boys % Positive Response			Girls % Positive Response		
	1935	1953	1959	1935	1953	1959
Activities						
I watched athletic games	65	67*	62*	24	46	41
I played athletic games	77	80*	81*	28	79	57
Occupations						
Scout leader[a]	44	20	12	2	12	19
Things to Own						
Baseball and bat	78	82*	84*	20	67	43
Things to Do						
Be on a school team	74	75*	74*	34	63	41
Go to a playground	74	59*	57*	36	55	43*

[a] Differences are significant for the two sexes but in the opposite direction.

Smith (in press) and Rosenberg and Sutton-Smith (1959) of children's interests in games has compared the responses of boys and girls in this generation with the responses of each sex to similar questionnaire items in the 1920's (Terman, 1926, 1936). The authors report results for the fourth, fifth, and sixth grades which indicate that there has been a shift in the direction of increasing preferences of girls for what were formerly boys' games while boys tend to select fewer items which significantly differentiate their choices from those of girls. The authors interpret their findings as an indication that the female role perception is expanding in the direction of greater activity and increased masculinity. This tendency is shown in the present study for the responses listed in Table 8. Not only preferences but recent participation is reported (see also Table 9).

The question has been raised as to what the effect on the masculine role will be, if girls' interests are expanded to include traditionally masculine activities (Hartley, 1959a, 1959b). We might conjecture that, for this sample at least, boys may be staking out a new claim in the borderland of social respectability involving betting and gambling (Table 7). On the other hand, the increased heterosexual orientation of ninth-grade boys suggests that they may not be averse to sharing interests in some areas with girls.

The overall picture indicates no striking changes in the number of items eliciting sex differences within each temporal sample. Comparing the significant changes for each sex for the two samples (1935–1953) we find boys' responses showing greater changes over time, with occupational rejections by the recent sample accounting for much of this disparity (Table 9). Girls give relatively more positive total responses in the later sample.

Tyler (1955) has suggested that aware-

TABLE 9
SEX DIFFERENCES: SIGNIFICANT DIFFERENCES IN PREFERENCES, ETC. 1935–53
(Test items with significant differences)

Test Category	No. of Items in Test	Boys		Girls	
		Increased Acceptance	Decreased Acceptance	Increased Acceptance	Decreased Acceptance
Things I did last week	20	1	1	8	0
Where I went	16	0	2	5	1
Things I read	18	2	3	6	1
Things to do	23	3	4	6	1
Occupations	47	0	28	2	9
Things you talk about	26	9	3	9	0
Places to go	23	1	14	3	6
Things to own	17	1	12	2	2
Magazines to read	17	0	10	1	6
My opinions	35	15	2	11	2
Totals	242	32	79	53	28
		12.5%	33%	21%	12%

Note.—For boys, 111 items (45% of total) showed significant differences; for girls, 81 items (33% of total) showed significant differences.

ness of role "enables the individual to rule out activities and goals not in keeping with it." In her longitudinal sample, *Ss* disliked more things as they grew older. Our own longitudinal results support her findings to some extent in the case of girls but not in the case of boys. In the area of vocational choices, for example, girls decreased their likes significantly on 20% of the items and increased them on 8%, from the 6th through the 12th grade. Boys, on the other hand, decreased their likes significantly on only 4% of the items, and increased them on 15%.

The maturity status of ninth-grade girls, higher than that of their male classmates, may be a factor in these findings. The nature of the items, the age of the *Ss*, and cultural differences, would also be determining factors in a comparison with other studies.

In relation to psychological sex-role development the data are not definitive. There is some blending of sex-role patterns but only as it is expressed concomitantly with earlier social maturing. There is some expansion of feminine interests into the male domain, but this is principally seen in sports activities. In the area of vocational preferences, girls' interests seem to be oriented toward supplementing the family income rather than toward a career. There is some constriction of the boys' vocational choices but this may be because we are comparing good times with the earlier depression era.

Members of the early growth study are now adults, some with adolescent offspring of their own. In interviews they often express their opinions about today's young people in contrast to their own remembered youth. They do not need to be told that adolescents now are more mature socially than in their day, though they may exaggerate the differences. In response to an interview question in this area, one of these young men said:

> I was early-maturing but not by today's standards. I think part of it is the way they dress now. The girls wear lipstick and look older. If they had dressed that way when I was a kid, I would have been sure they would bite me.

There may be some relationship between the frequent complaint of these study members in interviews: "Why didn't someone make me study more when we were in school?" and our findings that ninth graders say they *do* study more now. Strange as it may seem, one young man's statement: "I didn't know that the reason I was getting poor grades was that I didn't study" is not an isolated remark. Having become aware of this sequential relationship, such individuals now apply parental pressures for "more studying."

So in a sense, the present attitudes of our original sample, and their contemporaries, now parents and citizens, are among the environmental factors affecting changes in the directions which we have seen. Parental influences, evidence of earlier maturing (Meleny, 1959; Tanner, 1955) altered school programs, historical events, all give direction to the secular trends in interests and attitudes recorded in this study.

Summary

This report compares the activities, interests and attitudes of three ninth-grade classes in the same school but separated in time by 18 and 24 years (1935, 1953, 1959). The samples are closely similar in age and IQ distributions, and fairly comparable in a number of social measures. The earlier group, members of the Oakland Growth Study, provided age-trend data through tests given at regular intervals over a seven-year period.

The more recent generations of ninth graders indicate greater maturity of heterosexual interests, more serious purpose and a more tolerant attitude toward social issues. They reject more "childish" activities.

Alterations over time in sex-role development are not striking. There is some blending of sex-role patterns in the more

recent samples as expressed concomitantly with earlier social maturing. Currently, girls' interests have expanded in the area of sports activities. The responses of boys in the '50s indicate rejection of a greater number of occupations. Perhaps this is a reflection of occupational preferences as expressed during full employment as contrasted with the depression years. A greater verbal acceptance of "careers for women" is not accompanied by eagerness for a career among ninth-grade girls in the present generation. A variety of influences must be considered in accounting for the changes in interests and attitudes over the period of the study.

REFERENCES

EDWARDS, A. M. A social reference: Economic grouping of the gainful workers of the United States. *J. Amer. Statist. Ass.*, 1933, **28,** 377–387.

FURFEY, P. H. The measurement of development age. *Amer. Educ. Res. Bull.*, 1928, **2**(10), 40.

HARRIS, D. B. Sex differences in the life problems and interests of adolescents 1935–1957. *Child Develpm.*, 1959, **30,** 453–460.

HARTLEY, RUTH E. Sex-role pressures and socialization of the male child. *Psychol. Rep.*, 1959, **5,** 457–468. (a)

HARTLEY, RUTH E. Some implications of current changes in sex-role patterns. Based on address, New York Society of Clinical Psychologists, 1959. (b) (Mimeo.)

JACOB, P. E. *Changing values in college: An exploratory study of the impact of college teaching.* New York: Harper, 1957.

JONES, M. C. *The I.C.W. Interest Record.* Inst. Child Welf., Univer. California, 1940.

LEHMAN, H. C., & WITTY, P. A. *The psychology of play activities.* New York: A. B. Barnes, 1927.

MARSHALL, M. M. Awareness of sex-role and children's interest. Unpublished doctoral dissertation, Univer. California Education Library, 1958.

MELENY, H. E. Motor performance of adolescent girls today as compared with those of twenty-four years ago. MA thesis, Univer. California, 1959.

PRESSEY, S. L. Changes from 1923 to 1943 in the attitudes of public school and university students. *J. Psychol.*, 1946, **21,** 173–188.

PRESSEY, S. L., & PRESSEY, L. C. Development of the Interest-Attitude Tests. *J. appl. Psychol.*, 1933, **17,** 1–16.

REMMERS, H. H., & RADLER, H. W. *The American teen-ager.* Bobbs-Merrill, 1957.

ROSENBERG, B. G., & SUTTON-SMITH, B. The measurement of masculinity and femininity in children. *Child Develpm.*, 1959, **30,** 373–380.

ROSENBERG, B. G., & SUTTON-SMITH, B. A revised conception of masculine-feminine differences in play activities. *J. genet. Psychol.* (in press).

SANFORD, N. (Ed.) Personality development during the college years. *J. soc. Issues*, 1956, **4,** 12.

STOLZ, H. R., JONES, M. C., & CHAFFEY, J. The junior high school age. *Univer. High Sch. J.*, 1937, **2,** 15.

SYMONDS, P. S. A studiousness questionnaire. *J. educ. Psychol.* 1928, **19,** 152–167.

SYMONDS, P. S. Life interests and problems of adolescents. *Sch. Rev.*, 1936, **44,** 506–518.

TANNER, J. M. *Growth at adolescence.* Springfield, Illinois: Charles C Thomas, 1955.

TERMAN, L. *Genetic studies of genius.* Vol. I. Palo Alto: Stanford Univer. Press, 1926.

TERMAN, L., & MILES, C. C. *Sex and personality.* New York: McGraw-Hill, 1936.

TRYON, R. C. *Identification of social areas by cluster analysis.* (A general method with an application to the San Francisco Bay area) Berkeley: Univer. California Press, 1955.

TRYON, R. C. Identity of attitude and demographic dimensions that distinguish social areas. *Amer. Psychologist* 1959, **14,** 371. (Abstract)

TYLER, LEONA E. The organization of likes and dislikes in ten year old children. *J. genet. Psychol.* 1955, **86,** 33–44.

WALKER, H. M., & LEV, J. *Statistical inferences.* New York: Holt, 1953.

WEBSTER, H. Changes in attitudes during college. *J. educ. Psychol.* 1958, **49,** 109–117.

(Received March 17, 1960)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 187–190

SHORT-FORMS OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN IN THE EDUCABLE, NON-INSTITUTIONALIZED, MENTALLY RETARDED[1]

LEWIS SCHWARTZ
Indianapolis Public School System
AND EUGENE E. LEVITT
Indiana University Medical Center

Although many attempts have been made to develop abbreviated scales or short-forms of the Wechsler-Bellevue and the WAIS, there have been only a few studies of this type dealing with the WISC. Geuting (1959) reported the most valid (i.e., having the highest correlation with the Full Scale score) triads and tetrads of the WISC at the 7½, 10½, and 13½ year-old levels, based on Wechsler's own standardization data. The short-form validities were almost all between .90 and .95. On the other hand, Yalowitz and Armstrong (1955) found poor short-form validities for a heterogeneous child guidance clinic sample ranging in age from 5 to 15 years. Their highest coefficient was only .61. Carleton and Stacey (1954), using a sample of "suspected" mental defectives and emotionally disturbed children, obtained validities ranging from .64 to .88 for selected short-forms. A study of short-forms of the WISC with institutionalized mentally retarded boys was recently completed by Smith (1959). His report furnished data from which the correlation of any combination of subtests can be derived by means of multiple regression equations. The average IQ in the sample was 75, but the range indicated that about 40% of Smith's subjects would probably not be classifiable as mentally retarded.

The present report concerns a comprehensive investigation of a large number of short-forms of the WISC with an educable, mentally retarded, noninstitutionalized group of children. It grew out of the need for developing short-forms for quick re-evaluation of children in special classes for the mentally retarded in the Indianapolis public school system. It provides information whereby the composition of WISC short-forms which are useful for this purpose may be obtained.

Subjects and Procedure

The sample consisted of an unselected group of 177 children tested by the same psychologist and found to be eligible for placement in special classes for the educable, mentally retarded in the Indianapolis school system. There were 112 boys and 67 girls. Sixty-eight percent of the children were Negro, and 32% were white. About 90% of the children came from schools in lower and lower middle-class neighborhoods. The age range was from 9 years and 10 months to 15 years and 11 months with a mean of 13 years and 4 months. The range of Full Scale intelligence quotients on the WISC was 50 to 80 with a mean of 70.

The customary 10 subtests of the WISC were administered to 145 of the children. In the 32 instances where one subtest was omitted, an average value was assigned.

Statistical analyses of data were carried out on IBM equipment by the Research Computing Center of Indiana University. Correlations between the Full

[1] This study was supported by a grant from the Indianapolis Foundation of Indianapolis, Indiana, with the approval of Herman Shibler, then general superintendent of education, Indianapolis Public Schools.

Scale score on the one hand, and each individual subtest, and the sum of subtest scores for all possible duads, triads, tetrads, pentads, and hexads of subtests were computed. There were 45 duads, 120 triads, 210 tetrads, 252 pentads, and 210 hexads, making a total of 847 coefficients. Correlations for combinations of subtests greater than 6 were not computed since the economy in using a short-form of more than 6 subtests instead of the Full Scale would be too small.

Results

The correlations of individual subtests with the Full Scale range from .617 for PA[2] to .438 for CO. They are generally .10 to .15 correlation units below the comparable correlations reported for the standardization sample (Wechsler, 1949), but the relative magnitudes are similar. The rank order correlation of size of the coefficient between our results and the standardization data is .60, $p = .05$.

The customary procedure in a study of short-forms is to present the group of duads, triads, etc. having the highest correlations with the Full Scale. This does not seem to be a feasible approach in this paper. The range of coefficients of short-forms of each type is restricted. For example, the largest hexad coefficient is .945, while the 25th largest hexad coefficient is .932. The largest pentad coefficient is .920 and the 25th largest is .903, and so on. The differences tend to become more pronounced with smaller numbers of subtests—the range of the 25 largest duad coefficients is .11—but they are still relatively trivial.

Considerations other than validity are often of consequence in choosing a short-form. Administration time, handicaps in the child, or the examiner's personal preferences for particular subtests may dictate the choice. When the range of validity coefficients is limited, these considerations can be allowed heavier weight. It would therefore be misleading to present any small group of short-forms of those having the highest correlations with the Full Scale, which would imply that these are recommended for use by the school psychologist. Rather, we have elected to report a selection of five of each type of short-form, chosen from among the top two dozen correlations with the Full Scale. An attempt was made to include short-forms which varied maximally in composition. These will serve simply to give some idea of the disparity in coefficient size among the various types of subtests, and between our results and those of previous investigators. A more detailed account of our findings is on file and can be obtained by interested individuals.[3]

The 25 short-forms which we have selected are listed in Table 1 along with the respective correlations with the Full Scale score, and regression equations for predicting the Full Scale score from the short-form in scaled score (not IQ) units. The IQ can then be determined from Table XI B of the WISC *Manual* (Wechsler, 1949).

In general, the short-form validities we obtained are somewhat lower than those found by Geuting (1959) for normal children, and slightly lower than those reported by Carleton and Stacey (1954) for their mixed sample. This is about what one would expect; the correlations are largest when the sample is most heterogeneous,

[2] Abbreviations for subtests used in this report are conventional: I (Information), C (Comprehension), A (Arithmetic), V (Vocabulary), S (Similarities), PA (Picture Arrangement), PC (Picture Completion), BD (Block Design), OA (Object Assembly), and C (Coding).

[3] Tables listing the most valid 10 duads, 15 triads, 15 tetrads, 26 pentads, and 28 hexads, with a regression equation for each, have been deposited with the American Documentation Institute. Order Number 6250, from ADI Auxilliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm, $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

and tend to become attenuated as the sample becomes more homogeneous.

Discussion

There is no completely objective way to establish the entitlements of an adequate short-form of an intelligence test. The estimated magnitude of the error made in using a short-form instead of the Full Scale depends upon the variability of Full Scale scores, the correlation between them and the short-form scores, and the level of confidence with which we wish to make the predictions. In this study, the *SD* of the Full Scale scores is 10.04 in scaled score units. Using the customary .05 level of confidence, the estimated error when the correlation is .95 is ±6.1 scaled score units. For a correlation of .90, it is 8.6 scaled score units, for a correlation of .85, 10.4 units, and for a correlation of .80, 11.8 units.

The applied user of the short-form is, of course, free to decide for himself how much risk he is willing to take. But it would certainly seem that a potential error of 10 scaled score points is too high a price to pay for economy. If one hence insists on a correlation of .90 or higher, none of the duads and triads, and only a few tetrads, are useable. Only pentads and hexads would qualify.

The applied use of our results should properly be restricted to children of approximately the same description as those in our sample. This means that the tester must have some a priori knowledge of the child's intellectual level, which probably precludes the use of any short-form for an original intellectual evaluation. The purpose of the development of these short-forms was to furnish a method of re-evaluating the intelligence of known mentally retarded children, as in school systems where this must be done periodically. Such groups of children in most metropolitan areas and large towns will probably conform to the specifications of our sample.

TABLE 1

A Sampling of the Most Valid WISC Short-Forms of Each Type

Subtests	Correlation with Full Scale Score	Regression Equation[a] (Y =)
A PA	.789	2.5X + 31.05
S PA	.742	2.27X + 32.88
V BD	.741	2.86X + 25.72
PA OA	.722	2.05X + 35.42
A V	.717	2.09X + 28.35
A V PA	.843	2.09X + 23.00
A PA OA	.834	1.96X + 25.66
S V BD	.823	2.37X + 17.60
V PA BD	.817	1.97X + 24.66
C A PA	.812	2.00X + 23.41
I A PA OA	.893	1.87X + 18.05
A S PA BD	.879	1.74X + 19.50
V PC PA CO	.878	1.74X + 16.14
C A PA BD	.876	1.78X + 17.15
S PC PA CO	.872	1.69X + 17.74
A V PC PA CO	.920	1.55X + 12.32
I A S PA OA	.916	1.61X + 14.37
A S V PA BD	.915	1.57X + 13.86
S V PC BD CO	.913	1.69X + 7.68
V PC PA BD CO	.905	1.50X + 13.50
S V PC PA BD CO	.945	1.39X + 8.72
A S V PC PA CO	.943	1.39X + 9.04
C A PC PA OA CO	.943	1.39X + 8.19
I C A PA BD OA	.938	1.45X + 9.37
C S V PC BD CO	.936	1.53X + 2.46

[a] To obtain the estimated Full Scale score predicted by a short-form, substitute the *S*'s short-form score for X in the regression equation and solve for Y (the predicted Full Scale score). This gives the Full Scale score in scaled score units. The IQ can then be determined from Table XI on page 26 of the WISC *Manual* (1949).

Summary

This study was an attempt to develop short-forms of the Wechsler Intelligence Scale for Children for use in re-evaluating educable, noninstitutionalized, mentally retarded children. Correlations were obtained between the Full Scale WISC score and the individual subtests, and all possible duads, triads, tetrads, pentads, and hexads of subtests for a group of such children. The statistical computations suggest that only pentad and hexad short-forms can be substituted for the Full Scale with relatively small loss in accuracy of prediction.

REFERENCES

Carleton, F., & Stacey, C. L. Evaluation of selected short forms of the Wechsler

Intelligence Scale for Children (WISC). *J. clin. Psychol.*, 1954, **10**, 248–261.

GEUTING, MARY P. Validities of abbreviated scales of the Wechsler Intelligence Scale for Children. Unpublished master's thesis, Fordham Univer., 1959.

SMITH, E. A. A normative study of the Wechsler Intelligence Scale for Children with high-grade retarded boys. Unpublished paper, Wayne County Training School, Northville, Mich., 1959.

WECHSLER, D. *Manual for the Wechsler Intelligence Scale for Children*. New York: Psychol. Corp., 1949.

YALOWITZ, J. M., & ARMSTRONG, RENATE G. Validities of short-forms of the Wechsler Intelligence Scale for Children (WISC). *J. clin. Psychol.*, 1955, **11**, 275–277.

(Received October 1, 1959)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 191-194

GRADE DIFFERENCES IN TRANSFER AS A FUNCTION OF THE TIME INTERVAL BETWEEN LEARNING TASKS

MICHAEL GLADIS[1]

Johnstown College, University of Pittsburgh

The purpose of this study was to investigate transfer in third-, fifth-, and seventh-grade children in a verbal learning task when the time interval between Task I and Task II was varied. As such, it was stimulated by the incomplete nature of the evidence concerning changes in transfer as a function of age and time. Generalizations can be found in various textbooks stating that transfer increases with age in children but there appears to be little supporting evidence as to the nature and kinds of increases. In regards to the variable of the time interval between the first and second learning tasks, the evidence is such that it has led McGeoch and Irion (1956) to state that

> there are many problems connected with the retention of transfer effects which need to be worked out. In particular, the relationship between time and transfer needs to be investigated over a considerably wider range of learning situations. (McGeoch & Irion, p. 341).

METHOD

Subjects. Three groups of 45 *Ss* each were taken from the third, fifth, and seventh grades. The fifth-grade students represented the entire class of students at the Southmont Grammar School.[2] The seventh-grade children were selected randomly from some 130 students in this grade at the same school. Since there were only 38 third-grade students enrolled in the Southmont School, an additional seven were obtained from the Tioga Street School in Westmont, which is a neighboring suburb. While in each grade level there were several children younger and older than the norm age, the average age for each group closely approximated 8, 10, and 12 years.

Design. Ten *Ss* from each grade level were assigned randomly to one of three experimental conditions. The three experimental conditions were based on differences in the time intervals between the learning of Task I and Task II. The three time intervals were 5 minutes, 2 days, and 14 days. These groups will be designated as 0, 2, and 14 respectively. The remaining 15 *Ss* from each grade level comprised the control group for that grade.

Learning materials and apparatus. The learning material consisted of five pairs of one syllable four letter nouns. The words were selected from a second-grade reading list to rule out problems of unfamiliarity and reading difficulty. The second learning task contained the same response words as in the first list but the stimulus words were changed. Therefore, the transfer design was of the A-B, C-B type. The learning material was presented on a Hull-type memory drum. Each stimulus word appeared alone for 4 sec. and then the stimulus and response words appeared together for 4 sec. The intertrial interval was 8 sec. Five different orders of the lists were used to control for serial effects.

Procedure. After each experimental *S* was given the usual instructions for paired-associate learning, they learned a practice list which consisted of 3 pairs of one syllable nouns. They then learned Task I. Learning on Task I was carried out to one

[1] Now at Veterans Administration Hospital, Leech Farm Road, Pittsburgh, Pennsylvania.

[2] The author wishes to express his gratitude to Thomas Kochuba, Principal of the Southmont Grammar School, for his interest and invaluable assistance throughout the course of the investigation as well as to the officials and teachers of the Southmont and Tioga Street Schools for their cooperation. He also wishes to thank A. W. Bendig for his advice on statistical procedures.

trial beyond the first perfect repetition. Group 0 was given a 5-min. rest period during which they were permitted to browse in the library where the experiment was conducted. They then relearned the practice task and went on to learn Task II. The criterion of learning for Task II was one perfect repetition. The same procedure was utilized with Groups 2 and 14 except that the time interval between Task I and Task II was 2 and 14 days respectively. The control groups learned the practice task and then immediately learned Task II to a criterion of one perfect repetition.

Results

Learning on Task I. The mean number of trials to the criterion on Task I for the third-, fifth- and seventh-grade experimental groups was respectively 16.2, 9.4, and 10.3. An analysis of variance of the learning scores on Task I indicated that there was a significant difference between the grade group means at the .01 level. Separate *t* tests indicated that the mean number of trials for the third-grade group was significantly greater than the fifth- and seventh-grade means, but the difference between the fifth- and seventh-grade means was not significant. The mean number of trials to the criterion for the time interval groups did not differ significantly.

Learning on Task II. The mean number of trials to the criterion on Task II for each of the experimental and control groups is presented in Table 1. The mean

TABLE 1
Mean Number of Trials to the Criterion for the Experimental and Control Groups in Learning Task II

Grade	Time Interval 0	2	14	Composite	Control Group
Third	12.9	13.1	14.0	13.0	10.7
Fifth	7.4	8.6	7.0	7.7	7.6
Seventh	4.8	9.4	6.7	7.0	8.5
Composite	8.4	10.4	9.2		

TABLE 2
Mean Transfer Scores for Each of the Experimental Subgroups

Grade	Time Interval 0	2	14	Composite
Third	−3.3	−5.5	−10.5	−6.4
Fifth	8.7	4.4	7.6	6.9
Seventh	20.6	2.6	24.4	18.9
Composite	11.7	0.5	7.2	

scores for the third-, fifth- and seventh-grade experimental groups were 13.3, 7.7, and 7.0. An analysis of variance of the learning scores on Task II indicated that there was a significant difference between the grade group means at the .01 level. Separate *t* tests indicated that the mean number of trials for both the fifth and seventh grades was significantly less than the third-grade mean, but again the difference between the fifth- and seventh-grade means was not significant. The time interval group means also did not differ significantly.

The control group means for the third, fifth, and seventh grades respectively were 10.7, 7.6, and 8.5. While the trend of these mean scores is similar to that of the experimental groups on Task I and II, they did not differ significantly at the .05 level.

Transfer effects. To obtain an estimate of transfer for each *S* the formula

$$\frac{\text{Mean score of the Control group}^{a} - \text{Experimental } S\text{'s score}}{\text{Mean score of the Control group} + \text{Experimental } S\text{'s score}}$$

was utilized. This formula was used since it has the major advantage that the limit for both positive and negative transfer is 100% and is derived from one recommended by Murdoch (1957).

Table 3 presents a summary of the analysis of variance of the transfer scores.

[a] Each grade group has its own control group.

A difference significant at the .01 level was obtained between the grade group means which were −6.4, 6.9, and 18.9 for the third, fifth, and seventh grades respectively. Separate *t* tests indicated that both the fifth- and seventh-grade mean transfer scores differed significantly at less than the .05 level from the third-grade mean, but the difference between the fifth- and seventh-grade means did not reach the .05 level of significance. Furthermore, the analysis indicated that the time interval groups did not differ significantly, nor was the interaction between grade level and time interval significant.

TABLE 3
ANALYSIS OF VARIANCE OF TRANSFER SCORES

Source	*df*	*MS*	*F*
Between Groups	8	1760.8	2.76*
Grades (G)	2	4906.0	7.53**
Time Interval (T)	2	948.5	1.49
G × T	4	645.0	1.01
Within Groups	81	638.0	
Total	89		

* $P < .05$.
** $P < .01$.

DISCUSSION

The major finding of this study was that a significant difference in amount of transfer was obtained among the three grade level subjects used in this study. The trend of the changes in transfer among the three grade levels suggests that the kind and amount of transfer obtained may be a function of developmental factors, since the mean transfer score for the third-grade group was negative while it was positive for the fifth- and seventh-grade groups, with the seventh grade showing more transfer than the fifth-grade group. An individual analysis of the data indicated that 18 out of the 30 third-grade *Ss* showed negative transfer while only 13 of the fifth-grade *Ss* and 6 of the seventh-grade group evidenced negative transfer. It is interesting to note the occurrence of negative transfer in view of the fact that the transfer design, A-B, C-B, was one in which the prediction from Osgood's surface (1949) would be that positive transfer would occur. The results of this study suggest that the generalization that positive transfer will occur when the responses are identical but the stimuli change from Task I to Task II might not be a valid one for all age groups.

The failure to obtain any significant differences among the various time intervals used between Task I and Task II is consistent with findings by Bunch (1938) and Bunch & McGraven (1936) who, using college students as *Ss*, also failed to obtain any significant differences. It may be that the time intervals used in this study, 5 min., 2 days, and 14 days, which were similar to those used by the above-mentioned investigators, are too long to pick up any effects, since Hamilton (1950) found that the number of correct anticipations tended to decrease as the length of the time interval between Tasks I and II increased from 8 sec. to one hour but showed little change for longer periods of time. It should be noted that in the present study the least amount of transfer occurred with the 2-day group whose mean transfer score was only .5 in contrast to the mean scores of 11.7 and 7.2 made by the 0 and 14-day groups. Further studies using different time intervals appear necessary to clarify the issue.

A final point is that a larger difference in the rate of learning was found between the third- and fifth-grade groups than between the fifth- and seventh-grade groups. The rate of learning on Tasks I and II for the third-grade experimental group was significantly slower than the fifth- and seventh-grade experimental groups. Furthermore, the difference between the third-grade control group and the fifth-grade control group, while not quite reaching the .05 level of significance, was larger than the difference between the fifth- and seventh-grade control groups. Actually,

the mean score for the fifth-grade experimental group was less than the mean score of the seventh-grade experimental group on Task I, and the same trend was found between the fifth- and seventh-grade control group means. In general, then, all this seems to suggest that the difficulty level of the learning tasks used in this study is such that peak performance is probably attained by the time the children have reached the fifth grade.

Summary

This study investigated transfer in third-, fifth-, and seventh-grade *Ss* in a verbal learning task when the time interval between Task I and Task II was varied through the three conditions of 5 min., 2 days, and 14 days. The learning material consisted of meaningful paired associates and the transfer paradigm was of the A-B, C-B type. A significant difference in the amount of transfer was obtained between the three grade groups. The mean transfer scores for the fifth- and seventh-grade groups were positive and significantly different from the mean transfer score of the third-grade group which was negative. No significant differences were obtained between the different time intervals. The results were discussed in terms of their implication for transfer theory and related to previous findings in this area.

REFERENCES

Bunch, M. E., The amount of transfer in rational learning as a function of time. *J. comp. Psychol.*, 1938, **25**, 481–496.

Bunch, M. E., & McGraven, V. G., The temporal course of transfer in the learning of memory material. *J. comp. Psychol.*, 1936, **22**, 325–337.

Hamilton, C. C., The relationship between length of interval separating two learning tasks and performance on the second task. *J. exp. Psychol.*, 1950, **40**, 613–621.

McGeoch, J. A., & Irion, A. L., *The psychology of human learning.* (2nd ed.) New York: Longmans, Green. 1956.

Murdoch, B. B., Jr., Transfer designs and formulas. *Psychol. Bull.*, 1957, **54**, 313–326.

Osgood, C. E., The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.

(Received October 2, 1959)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 195-198

STABILITY OF ITEM STATISTICS IN TWO INSTITUTIONS

ERNEST R. HILGARD
Stanford University

AND EARL L. JANDRON
San Jose State College

Although the theory of test construction has now reached an advanced stage (e.g., Gulliksen, 1950) course examinations in colleges and universities, including those that consist of objective test items, often fall short of desirable test characteristics. Test theory alone will not produce good examinations; it is necessary to give the items careful local consideration if the test is to meet educational as well as measurement objectives. The present paper bears on this problem by presenting some experiences in the testing of the same items at two institutions in which arrangements for the course differ widely.

The course selected for study was introductory psychology. At San Jose State College the course is taught in small sections meeting three times a week for a semester, and the course with which we are concerned is limited to first-semester freshmen. At Stanford the course is taught in a large lecture meeting four times a week for a quarter, with a one-hour discussion period in small groups conducted by teaching assistants. Most of those who take the course are sophomores, although there are a number of upper classmen and a few freshmen. The purpose obviously was not to study the comparability of the outcomes of the courses at San Jose State College and at Stanford, but rather to see whether or not the same questions were appropriate for use under such different conditions of teaching. Both sets of classes used the same textbook as part of the course requirement, although the outside readings differed widely both in kind and amount.

PROCEDURES

Two examinations were prepared and used as final course examinations. Examination 1, consisting of 230 questions, was administered to 221 students at Stanford in December 1958. Examination 2, also consisting of 230 questions, was administered to 125 students at San Jose in January 1959. Of the 230 questions on each of these examinations, 115 were in common to the two examinations. All questions were of multiple-choice type, the correct answer being selected from four alternatives.

While questions were carefully reviewed and edited, they had not previously been tested. They had been constructed in order to add to an existing item pool, and this constraint upon them led to the inclusion of some specialized items that increased the average level of difficulty beyond the level the authors consider desirable. The examinations, taken as a unit, were reasonably satisfactory, despite the fact that there were many questions included that were rejected following the pretest. The reliabilities (Kuder-Richardson) were for Examination 1, .82, and for Examination 2, .65. The lower reliability for the San Jose group (Examination 2) is probably a function of item difficulty, the items being somewhat too hard for the freshmen students inexperienced in this kind of examination.

Item analysis led to two measures:

1. *Item difficulty.* The number of correct replies was determined for each question and expressed as a per cent. This served as a measure of item difficulty, the higher per cent being, of course, the easier items.

2. *Discrimination index.* A discrimination index was computed for each item. This is a tetrachoric correlation between the single item and the total score on the examination. It is computed by determining the percentage successful on the item in each of two fractions of the whole examination, the upper 27% and the lower 27%. Then the tetrachoric correlation is looked up in an item analysis table (Fan, 1952).

Our results concern the stability of these two measures between the two college populations, and the relationships between the two measures.

Consistency of Relative Item Difficulty

Many differences in student selection, in stage of academic career, in time devoted to individual topics, contributed to a different average performance on the questions in common at the two teaching centers, but this is an advantage for the purposes of our analysis. At San Jose the mean per cent passing each of the 115 questions was 47.6, with a standard deviation of 20.6; at Stanford the mean per cent passing was 63.5, with a standard deviation of 19.2. This yields a highly significant difference between means ($CR = 10.7$). The distributions are of the questions according to their difficulty; a related, but different, statistic would be found for differences in mean scores for the students at the two places.

Despite the differences in average difficulty level, the *relative* difficulty of items at the two places was similar. The dispersions were similar, and the two distributions of difficulty were correlated to the extent of $r = +.80$. These results confirm an earlier report in which a correlation of .87 was found between item difficulty at the University of Michigan and at Rutgers University for the same 80 items used at both places (Hilgard, Teevan, & Wells, 1954.) This relative consistency means that one college or university can profit by using items pretested at another, even though local circumstances shift the difficulty of the items up or down.

Consistency in Item Discrimination

Item discrimination, as measured by tetrachoric correlations between the passing and failing of individual items and the success on the test as a whole, turned out to be not significantly different at the two places. The mean tetrachoric for the 115 questions at San Jose was .26 with a standard deviation of .18, and at Stanford .25, standard deviation .15. The product-moment coefficient of correlation for the discrimination turns out to be +.31 which indicates that test items that discriminated at one place tended to discriminate at the other. However, the correlation is rather low and suggests that an instructor can not assume that because an item is very discriminating at one place it will invariably be very discriminating at another.

There are some problems that arise when tetrachoric correlations are used in item testing. In general, more difficult items yield lower correlations with the criterion, because the guessing factor becomes more important. The correction proposed by Carroll (1945) is of little help when it comes to item selection, for the lower reliabilities due to guessing do in fact persist in determining the actual test scores. Correlations with the total score also tend to be reduced for items that depart widely from the average difficulty of the items. Because of the correlation in item difficulty between the two tests, some correlation in discrimination indexes is to be expected.

Suggestions for Local Test Construction from Items Pretested Elsewhere

The comparability of the relative difficulty of items at the two places permits some suggestions as to the appropriate use

at one place of multiple-choice items pretested elsewhere. While the course under consideration was one in psychology, the suggestions should apply as well to other courses using multiple-choice questions.

1. Items found satisfactory elsewhere are likely to provide a useful pool, but the average difficulty level may shift.

2. Local records of percentage passing should be kept, so that item difficulty can be controlled according to the nature of the students and of the course taught.

3. Where the pool of questions is large enough for appropriate control of difficulty, it is desirable to narrow the range of difficulty of the items. Keeping difficulty nearly constant will improve both the reliability and the validity of the test according to Cronbach and Warrington (1952) and Lord (1952).

4. In selecting for difficulty it is better to err on the side of choosing questions that are too easy than those that are too hard. It has been shown by Lord (1952) that the optimum level of difficulty is somewhat easier than halfway between a chance level and 1.00. For a four-choice item, this means a level above the mean between .25 and 1.00, or something over .625. By this criterion, the level of difficulty of the items was about right for the Stanford students, but the test was too difficult for the San Jose students, resulting in lower reliability. A practical advantage of using somewhat easier questions is that the resulting distribution of scores is likely to be skewed somewhat to the lower end, thus, making for good discrimination at the point where decisions about passing and failing have to be made.

5. Because of the face-validity of edited and pretested questions, correlation with the criterion can be assumed provided the correct answer is given at a level well above chance. Hence, the laborious computation of discrimination indexes is unnecessary in selecting questions to remain in the local pool of items. Because of multiple objectives in many courses, too high a dependence upon internal consistency (i.e., correlation with total score) may actually defeat some of the objectives. In that case, a dependence on face-validity in terms of one or more of the course objectives, plus a high percentage passing, will conserve the good items.

Because all of these controls can be exercised merely through keeping a record of item difficulty, those who seriously wish to improve measurement in their courses are well advised to keep such records on the item pools from which their examinations are constructed. If they do this, they may use with advantage items pretested elsewhere.

Summary

1. Multiple-choice items to be used in courses in introductory psychology were pretested at San Jose College and at Stanford University, where the clientele and arrangements for the first course differ markedly. There were 115 questions in common between the two tests, making comparisons possible.

2. Despite a statistically significant difference in average difficulty of the items, difficulty correlated .80 between the two tests. The mean discrimination indexes (correlations of individual items with total test) did not differ significantly at the two centers, and the two sets correlated .31 with each other. The results for both item difficulty and discrimination suggest the usefulness at one place of items pretested elsewhere.

3. Because of the importance of item difficulty in constructing a satisfactory examination, local norms are important. It is preferable to keep item difficulty relatively uniform, and to err on the side of having items too easy rather than too difficult.

REFERENCES

Carroll, J. B. The effects of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, **10**, 1–19.

CRONBACH, L. J., & WARRINGTON, W. G. Efficiency of multiple-choice items as a function of spread of item difficulty. *Psychometrika*, 1952, **17**, 127–146.

FAN, C-T. *Item analysis table*. Princeton, N. J.: Educational Testing Service, 1952.

GULLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.

HILGARD, E. R., TEEVAN, R. C., & WELLS, W. D. *Instructors manual*. New York: Harcourt, Brace, 1954.

LORD, F. M. The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika*, 1952, **17**, 181–194.

(Received October 23, 1959)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 199-207

THE EFFECTS OF VERBAL MEDIATION ON THE MODIFICATION OF CHILDREN'S ATTITUDES[1]

SAMUEL H. OSIPOW[2]
Syracuse University

The purpose of the present experiment was to investigate the effect of verbal mediation on the acquisition of attitudes. The general hypothesis was that labels and their associated stimulus components mediate approach and avoidance responses. A secondary hypothesis was that initial preferences for stimuli influence the rate of learning new affective labels for these stimuli.

From the school's standpoint, these hypotheses appear to be extremely relevant. On the one hand, the teaching-learning process is mediated by the use of language and other symbols. On the other hand, the modification of attitudes is among the more important of the school's objectives. Among the many theories concerning the development and modification of attitudes, the one advanced here attempts to demonstrate that language contributes to attitude formation because of the affective meanings that are learned to verbal symbols, thereby relating one feature of the teaching-learning process to one of the school's objectives. In order that the present study might be still more applicable to the school situation the subjects used were from a population of fourth-grade pupils.

The rationale is based on the theoretical discussion of higher order learning presented by Dollard and Miller (1950). Labeling, i.e., the connection of verbal responses to stimuli, is considered a basic process in this theory. The associative stimulus components of the label evoke response-produced cues leading to the distinctive emotional or instrumental responses.

On this basis, preferential behavior is assumed to be mediated by a mechanism similar to the rg-sg mechanism described by Hull (1931). Briefly, a name is learned for an unnamed object. This name serves as a stimulus for a learned evaluative label (Osgood & Suci, 1955), that has acquired affective associations, i.e., meaning (Osgood, Suci & Tannenbaum, 1957). A label loaded with positive affective associations mediates an implicit verbal response of approach, resulting in increased preference when attached to previously neutral objects. Conversely, negatively loaded labels mediate avoidance responses when associated with previously neutral objects, resulting in lower preferences for, and avoidance of, the object.

Following the general laws of learning, it is assumed that whatever label is connected to a stimulus is strengthened with each reinforced trial, and will be retained as long as it continues to be reinforced. On the other hand, an inappropriate label, one not reinforced, will be extinguished. Reinforcement will occur if the label provides *S* a useful basis for the discrimination (or generalization) of responses, retroactively provides primary drive reduction, or inhibits responses which provide a learned drive.

The foregoing rationale implies that new verbal labels may be substituted for labels

[1] This paper is based upon a dissertation submitted to the Graduate School, Syracuse University, in partial fulfillment of the requirements for the PhD degree. The dissertation was written under the direction of Francis J. Di Vesta. The author would like to thank Wallace R. McAllister and George G. Thompson for their assistance.

Portions of the present paper were read at the American Psychological Association Convention in Cincinnati, Ohio, September 1959.

[2] Now in the United States Army, Abilene, Texas.

with strong response potential provided the new labels are made appropriate by reinforcement. Accordingly, rewards provided for associating approach labels with noxious objects result in the extinction of the connection between the avoidance label and the object. The avoidance label is replaced by an approach label, which, in turn, elicits approach responses. The process involved in learning to attach avoidance labels to initially approach stimuli is fundamentally the same. Via general transfer principles it is assumed that the rate of acquisition of a label for a previously unlabeled object is not a function of the meaning of the label since zero transfer is involved. On the other hand, the rate of learning labels for objects already possessing names is expected to be an inverse function of the amount of difference between the meanings of the original and new labels since differing degrees of positive and negative transfer are involved.

Several empirical studies in learning (Schaeffer & Gerjuoy, 1955; Spiker, 1956; Spiker, Gerjuoy, & Shepard, 1956) have demonstrated that labels aid in the development of new discriminations or generalizations. Parallel investigations (Stagner & Britton, 1949; Staats & Staats, 1958; Staats, Staats, & Biggs, 1958; Rhine & Silun, 1958) provide evidence that attitudes can be learned by conditioning techniques, while Eisman's (1955) study provided support for the application of mediated generalization to the study of attitudes.

The above rationale led to the following hypotheses:

1. The rate of acquisition of labels for unlabeled objects is not a function of the meaning of the labels.

2. The rate of acquisition of labels for labeled objects is an inverse function of the disparity between the meanings of the old and new labels.

3. A change in attitude is mediated by the preference associations stimulated by labels.

Method

Subjects

The *Ss* used in the experiment were 25 male and 31 female fourth-grade pupils (mean age 124.6 months, standard deviation 5.28 months). Seven additional *Ss* were discarded for failure to reach a learning criterion discussed below.

Materials

Eight sets of wooden blocks were used as stimuli in Phase C of the experiment. Each set was painted a different color and contained one cube, one three-dimensional flat triangle, one elongated cylinder, one short cylinder, and one parallelogram. Nonsense figures were made of black oaktag paper and used as stimuli in Phases B and D. The figures were ambiguous in that none of the familiar names (triangle, rectangle, and so on) were appropriate labels. A wooden board with three indentations was used in Phases B and C to prevent the presence of extraneous cues regarding the location of a paper token used as reinforcement and placed beneath the object identified as "correct."

Procedure

The procedure was divided into four phases, as summarized in Table 1. All four phases were conducted in one session. Phases B, C, and D were modeled after the mediational sequence (A-B, B-C, A-C) demonstrated (Bugelski & Scharlock, 1952) to be the most effective in mediating responses.

Which of the *Ss* was to take part in the experiment at any particular time was determined by the teacher, who chose the next *S* to see *E*. *E* then matched *S*'s name to a number (from one of four) selected at random to assign the *S* to a group. The last three *Ss* were assigned to groups lack-

ing an *S* in order to equalize the *N* in each cell of the design.

Phase A. Four pretests were administered before *S* participated in the experiment. Pretest 1 was designed to determine whether *S* knew the meanings of the words PRETTY and UGLY, the two "unconditioned stimuli." *E* showed Binet cards, Age IV-6, and asked which person on the card was prettier and which was uglier. In Pretest 2, *E* asked *S* which of the people depicted on the card he would rather be like in order to determine whether *S* had appropriate meanings for the words PRETTY and UGLY. Pretest 3 was designed to secure *S*'s initial preference ranks for the nonsense figures. The figures were placed before *S*, who was instructed to look at the pictures carefully and select the one he preferred. When the selection had been made, *E* removed the figure, and instructed *S* to select that one of the remaining figures he now preferred. This procedure (Schroder, 1956) was followed until all figures were chosen. A rank order preference for the nonsense figures for each *S* was thus obtained. Pretest 4 was designed to determine *S*'s preference for colors. The rank order of preferences for eight cubes, each a different color, was obtained by the same ranking procedure used in Pretest 3.

Phase B. In this phase *S*s were trained to label the nonsense figures. The general method required each *S* to attach a color name to one of his middle-ranked figures. The *S* was instructed that he was about to play a game in which he could earn tokens that could be traded for a toy. The board with three nonsense figures placed over the holes was presented to *S*. Two of the figures corresponded to those demanded by the experimental condition to which the *S* was assigned, the third was selected at random from the remaining figures. *S* was instructed that: (*a*) two of

TABLE 1

SUMMARY OF PROCEDURES

Group	Phase A	Phase B	Phase C	Phase D
I	Pretests: 1. Knowledge of PRETTY and UGLY. 2. Affective value for PRETTY and UGLY. 3. Rank nonsense figures. 4. Rank colored cubes.	Learn to call 5th (some *S*s 4th) preferred nonsense figure by most preferred color name, and 4th (some *S*s 5th) preferred nonsense figure by least preferred color name.	Learn to call most preferred color PRETTY and least preferred color UGLY.	Re-rank nonsense figures.
II	Same as Group I	Learn to call 4th (some *S*s 5th) preferred nonsense figure with 4th (some *S*s 5th) preferred color name, and 5th preferred figure (some *S*s 4th) with 5th (some *S*s 4th) preferred color name.	Learn to call 4th (some *S*s 5th) preferred color PRETTY and 5th (some *S*s 4th) preferred color UGLY	Same as Group I
III	Same as Group I	Same as Group I.	Learn to call most preferred color UGLY and least preferred color PRETTY	Same as Group I
IV	Same as Group I	Same as Goup II.	Learn to call 4th (some *S*s 5th) preferred Color X, and 5th (some *S*s 4th) preferred Color Y	Same as Group I

the figures had color names and one did not; (*b*) he was to determine what these names were; and (*c*) one of the figures was named Color 1 and the other figure Color 2 (the two appropriate color names). *E* then told *S* that he would be asked which figure was named Color 1. When he had decided, he was to name the figure, point to it, then lift it and look beneath it into the indentation. If he found a paper token he would know he was correct. If he did not find the token, then he was wrong, and the figure was not named Color 1. *E* then instructed *S* to close his eyes, and the procedure was begun. The selection of a figure for each of the two color names was considered one trial. The intertrial interval varied, being determined by the speed with which *E* set up the next presentation. The procedure was followed until *S* correctly answered five consecutive trials. After this criterion had been reached, 15 additional trials were administered. Seven *Ss* failed to reach the criterion after 65 trials and were discarded from the study. Two of these *Ss* were in Group I, three in Group II, one in Group III, and one in Group IV.

The specific treatment in Phase B required *Ss* in Groups I and III to learn to call one middle-ranked figure by the name of their most preferred color and the other middle-ranked figure by the name of their least preferred color. The order was determined at random so that some *Ss* learned to call their fourth preferred figure by the name of their least preferred color and their fifth preferred figure by the name of their most preferred color. The pairing of figure with color name was reversed for the alternate *Ss*. Groups II and IV learned to call one middle-ranked nonsense figure by the name of the other middle-ranked color.

Phase C. After *Ss* learned color names for nonsense figures they were trained to learn labels to selected colors of blocks. The general method was the same as that of Phase B except that the color names were replaced by PRETTY and UGLY for Groups I, II, and III, and by X and Y for Group IV.

The specific treatments were as follows: *Ss* in Group I named most preferred colors PRETTY and least preferred colors UGLY. *Ss* in Group II learned to name one middle-ranked color PRETTY and the other middle-ranked color UGLY. *Ss* in Group III learned to name most preferred colors UGLY and least preferred colors PRETTY. *Ss* in Group IV, a control group, learned to name one middle-ranked color X and the other middle-ranked color Y.

Phase D. After completion of Phase C, *E* informed *S* that enough tokens had been earned for a toy but that one more activity would be required before the prize could be chosen. *E* then presented the nonsense figures following the same procedure as in Pretest 3 of Phase A. Thus, Phase D consisted of a re-ranking by each *S* of his preferences for the nonsense figures.

RESULTS

The groups did not differ in the number of trials to criterion in Phases B and C. However, to permit a comparison between the rates of acquisition of the labels of the groups, the data were Vincentized by dividing the total number of trials to reach criterion for each *S* into fifths and computing the percentage of the number of correct trials for each block.

Phase B

The data are summarized graphically in Fig. 1 for all four groups during Phase B. Hypothesis 1 was tested by three specific predictions related to the performance of the four groups in Phase B.

Hypothesis 1a. The mean percentage of correct responses over blocks of trials will show an increase in performance for the groups in Phase B.

Hypothesis 1b. The mean percentage of correct responses over blocks of trials will

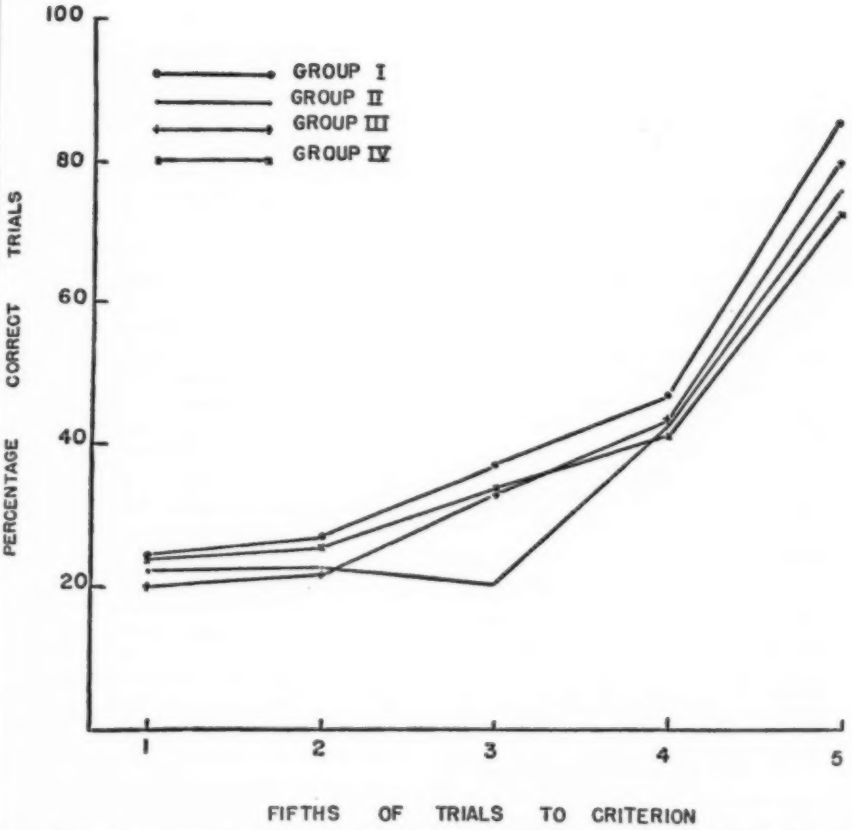

FIG. 1. Vincentized performance curves for Groups I, II, III, and IV during Phase B.

follow the same pattern for all groups in Phase B.

Hypothesis 1c. The mean percentage of correct responses over blocks of trials will be the same for all groups in Phase B.

It was necessary to test Hypothesis 1a to insure that the labels had been attached to the nonsense figures by the *Ss*. The data relating to this hypothesis were analyzed by a Type I analysis of variance (Lindquist, 1953). The summary of this analysis is presented in Table 2. The between trials F of 110.27 is significant ($p < .001$), thereby indicating improvement in performance.

Hypotheses 1b and 1c were of interest since differences in performance between the various groups in the nonsense figure labeling task might be related to differences in labeling colored objects and/or preferences for the figures. The interaction of trials by groups shown in Table 2 has an F ratio of 0.71, indicating that the differences between the slopes of the curves of the groups are small enough to have occurred by chance. The between groups F of 1.49, shown in Table 2, indicates that the vertical displacement of the curves did not differ in Phase B. Accordingly, Hypotheses 1b and 1c were not rejected.

TABLE 2
ANALYSIS OF VARIANCE OF PHASE B

Source	*df*	*MS*	*F*	*p*
Between Groups I, II, III, IV	3	702.37	1.49	>.05
Between *Ss* in the same grp.	52	471.26	—	—
Total between *Ss*	55	—	—	—
Between Trials I through V	4	29419.90	110.90	<.001
Interaction: Trials × Groups	12	190.17	0.71	>.05
Interaction: Pooled *Ss* × Trials	208	266.79	—	—
Total within *Ss*	224	—	—	—
Total	279	—	—	—

The conclusion is that no differences in performance existed among the groups during Phase B.

Phase C

The data for Phase C, in which *Ss* learned to attach the evaluative names to the colored blocks, were analyzed by analysis of variance. The specific hypotheses related to Hypothesis 2 concerned differences in performance among the groups within this phase.

The first of these subhypotheses was Hypothesis 2a, that the mean percentage of correct responses over blocks of trials in Phase C will be greatest for Group I, less for Group II, and least in Group III. The *F* ratio between groups is 6.25, indicating a significant ($p < .01$) difference among groups in the percentage of correct responses.

The possibility that the variances of the groups may have contributed to this difference was tested via Bartlett's test for homogeneity of variance. With 3 degrees of freedom and a chi square of 1.52, the null hypothesis that the groups comprised a population of common variance was not rejected.

Since the analysis of variance indicated that the differences among the groups were not likely to have occurred by chance, separate comparisons were made by the Scheffé test (1953) applied to the Vincentized means between groups for Phase C. The results of this analysis are summarized in Table 4. A significantly higher percentage of correct responses were made by Group I than by Group II ($p < .01$) or by Group III ($p < .05$). Groups II and III did not differ significantly in the mean percentage of correct responses ($p > .05$).

The second of the subhypotheses was Hypothesis 2b, that the mean percentage of correct responses over blocks of trials would show an increase in performance for the groups in Phase C. This hypothesis was of concern since the observation of mediation effects in Phase D depends upon the association of colors with evaluative

TABLE 3
ANALYSIS OF VARIANCE OF PHASE C

Source	*df*	*MS*	*F*	*p*
Between Groups I, II, III	2	12453.05	6.25	<.01
Between *Ss* in same group	39	1991.98	—	—
Total between *Ss*	41	—	—	—
Between Trials I through V	4	27019.54	41.83	<.001
Interaction: Trials × Groups	8	1680.71	2.60	<.05
Interaction: pooled *Ss* × Trials	156	645.87	—	—
Total within *Ss*	168	—	—	—
Total	209	—	—	—

TABLE 4
COMPARISON OF THE MEAN NUMBER OF CORRECT RESPONSES OVER TRIALS DURING PHASE C FOR GROUPS I, II, AND III

Mean	Contrast	Result	*p*
I 82.94	I minus II	I > II	<.01
II 56.70	II minus III	II = III	—
III 65.02	I minus III	I > III	<.05

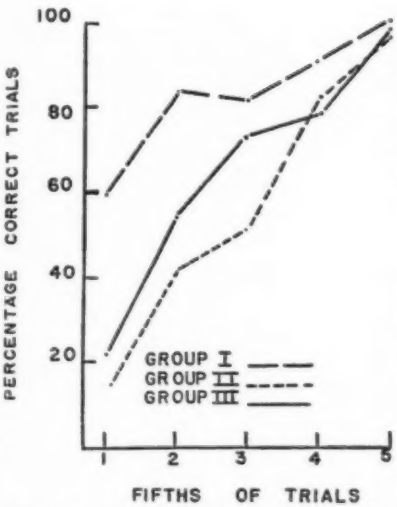

FIG. 2. Vincentized performance curves for Groups I, II, and III during Phase C.

names. The between trials F of 41.83 shown in Table 3 is significant ($p < .001$) and supports the hypothesis. The performance curves for Groups I, II, and III showing the mean percentage of correct trials per fifth of trials to criterion for Phase C are presented in Fig. 2.

The third of the subhypotheses in this phase was Hypothesis 2c. It was predicted that the mean percentage of correct responses over blocks of trials would follow a different pattern for each group in Phase C. In this hypothesis different performance curves were predicted for the experimental groups. The F of 2.60 for the between groups over trials interaction shown in Table 3 is significant ($.01 < p < .05$), supporting the hypothesis.

Phase D

The differences in rank between the nonsense figure that had been associated with the color called PRETTY and the figure associated with the color called UGLY were of interest in Phase D. For Group IV, the concern was with the differences in rank between the figures associated with x and y. The data relevant to these comparisons were analyzed by analysis of variance (Type I). Since the pairing of the fourth or fifth figure with the positive or negative stimulus was made at random, the initial rank of the figure was not considered in the analysis. The summary of the analysis is presented in Table 5.

The mediation hypothesis (Hypothesis 3) was tested through two subhypotheses by comparisons of the ranks of the figures and by comparisons between the ranks given by the four groups.

In Subhypothesis 3a it was predicted that the difference between the rank of the figure in the PRETTY sequence and the UGLY sequence would be significant for Groups I, II, and III. The between methods F of 9.21 (see Table 5) is significant ($p < .01$). The figures in the sequence in which the word PRETTY was used were ranked higher in preference than those in the sequence in which the word UGLY was used. These data support Hypothesis 3a.

In Subhypothesis 3b, it was predicted that the mean difference score of ranked preferences between the PRETTY and UGLY sequences would be the same for Groups I, II, and III, and that Group IV would

TABLE 5
ANALYSIS OF VARIANCE OF PHASE D

Source	*df*	*MS*	*F*	*p*
Between Groups I, II, III, IV(A)	3	1.38	0.33	>.05
Between *Ss* in same group	52	4.33	—	—
Total between *Ss*	55	—	—	—
Between Methods[a] I and and II[b] (B)	1	28.00	9.21	<.01
Interaction—A × B	3	0.24	0.08	>.05
Pooled *Ss* × B	52	3.04	—	—
Total within *Ss*	56	—	—	—
Total	111	—	—	—

[a] Method I refers to association with positive preference.

[b] Method II refers to association with negative preference.

have a smaller difference score than the other groups. The between Groups *F* of 0.32 shown in Table 5 indicates that no differences existed between the changes in rankings made by the four groups. Thus, Hypothesis 3b is not supported.

In order to investigate whether X and Y had positive and negative meanings which mediated the respective changes in Group IV, a follow-up study was conducted. All pupils in the classroom were given a paired comparison task to rate preferences for five letters, among which were the letters X and Y. An analysis of the number of times X and Y were chosen was performed by a *t* test for correlated data. The results of this analysis indicated that both the *Ss* and the other pupils in the classroom expressed more preference for Y than for X. This finding is the reverse of expectation, since in Phase D the figure called by the name of the color called X was more preferred while the figure called by the name of the color called Y was less preferred.

Discussion

All groups learned to label a nonsense figure with a color name at the same rate in the initial task. These results are consistent with the theoretical considerations involved in labeling presented earlier in this paper.

The differences in results obtained among the experimental groups' performance in attaching evaluative labels to color names can be accounted for by considering positive transfer effects. The first experimental group had learned labels consistent with its prexperimental attitudes, leading to the greatest amount of positive transfer, and superiority in performance. In the second and third experimental groups the labels did not agree, or were in lesser agreement, with prexperimental attitudes. This resulted in a smaller degree of positive transfer, and perhaps some negative transfer, from previous experience with the colors. This would account for their somewhat poorer performance than Group I. The lack of a difference in performance between Groups II and III indicates no differential effects in these groups in the connection of evaluative labels to the colored blocks.

The *Ss* in the three experimental groups changed preferences in the direction predicted. *Ss* ranked the nonsense figures higher in preference when they attached the evaluative label PRETTY to the color name and lower in preference when they attached the evaluative label UGLY to the color name. Thus, until the results of the control group are considered, the evidence corresponds closely with that of Eisman (1955) and Rhine and Silun (1958). The control group, it will be recalled, showed final attitudes similar to those of the experimental groups. The importance of the control group is now immediately apparent and cannot be overemphasized. When the results of the experimental groups are now compared with the control group it will be seen that the conclusion that meaning is the relevant factor in the mediational process must be tempered by the fact that similar findings occurred with the control group. This modified conclusion suggests that mediation may be a more complex phenomenon than has been suggested in the earlier literature. An accurate examination and identification of the relevant variables in the mediation of preferences and attitudes must necessarily depend upon more adequate controls than those used in the past. On the other hand, one cannot assume that the labeling process had no effect on preference. The most likely result of such an assumption would be no differences between the final ranks of the positive and negative nonsense figures. However, a difference was demonstrated in all groups, thus, the mediational process apparently has an influence on preference but not one specifically attributable to the meanings of the unconditioned stimuli.

Summary

The purpose of the present study was to investigate the effect of verbal mediation on the acquisition of preferences. Of secondary interest was the study of the rate of acquisition of labels under various transfer conditions. The theoretical basis of the study was Dollard and Miller's description of the cue-producing properties of labels.

Fifty-six fourth-grade children were divided into four groups, each group given different treatments. Groups I and III first learned to connect their most and least preferred color names to nonsense figures. Group I then labeled its more preferred color name with a positive evaluative name, while the procedure was reversed for Group III. Groups II and IV labeled nonsense figures with middle-ranked color names in the first task. Group II then labeled its middle-ranked color names with positive and negative evaluative words, while Group IV labeled its middle-ranked color names with X and Y. Finally, all groups were required to rank their preferences for the nonsense figures.

It was found that the groups did not differ in the rate of acquisition of color names for the nonsense figures. The rate of learning to name colors with evaluative words was related to the degree of transfer involved. The preferences for the nonsense figures were related to the positive or negative nature of the evaluative word in the mediational chain. However, Group IV, the control group, showed shifts equivalent to those of the three experimental groups. The results were interpreted as only partially supporting the mediation hypothesis in the formation of preferences. The importance of the control group in the demonstration of the role of meaning was indicated.

REFERENCES

Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.

Dollard, J., & Miller, N. S. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

Eisman, B. S. Attitude formation: The development of a color preference response through mediated generalization. *J. abnorm. soc. Psychol.*, 1955, **50**, 321–326.

Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.*, 1931, **38**, 487–506.

Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

Osgood, C. E., & Suci, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana, Ill: Univer. Illinois Press, 1957.

Rhine, R. J., & Silun B. A. Acquisition and change of a concept attitude as a function of consistency of reinforcement. *J. exp. Psychol.*, 1958, **55**, 524–529.

Schaeffer, M. S., & Gerjuoy, I. R. The effect of stimulus naming on the discrimination learning of kindergarten children. *Child Develpm.*, 1955, **26**, 231–240.

Scheffe, H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 87–104.

Schroder, H. M. Development and maintenance of the preference value of an object. *J. exp. Psychol.*, 1956, **51**, 139–141.

Spiker, C. C. Stimulus pretraining and subsequent performance in the delayed reaction experiment. *J. exp. Psychol.*, 1956, **62**, 107–111.

Spiker, C. C., Gerjuoy, I. R., & Shepard, W. O. Children's concept of middle-sizedness and performance on the intermediate size problem. *J. comp. physiol. Psychol.*, 1956, **49**, 416–419.

Staats, A. W., & Staats, C. K. Attitudes established by classical conditioning. *J. abnorm. soc. Psychol.*, 1958, **57**, 37–40.

Staats, A. W., Staats, C. K., & Biggs, D. A. Meaning of verbal stimuli changed by conditioning. *Amer. J. Psychol.*, 1958, **71**, 429–431.

Stagner, R., & Britton, R. H., Jr. The conditioning technique applied to a public opinion problem. *J. soc. Psychol.*, 1949, **29**, 103–111.

(Received November 4, 1959)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 208-212

SOCIABILITY AND ACADEMIC ACHIEVEMENT IN VARIOUS TYPES OF LEARNING SITUATIONS[1]

LESLIE R. BEACH
Whitworth College

In recent decades, educators have become increasingly concerned with the problem of improving college teaching. As a result, a veritable rash of formal and informal studies has cropped out in an effort to find the teaching methods which are the most effective in the college classroom.

Much of the formal research on college teaching has centered about differences in classroom instructional methods, with academic achievement being studied as the dependent variable (Birney & McKeachie, 1955). However, comparatively little attention has been given to the personality of the *learner* as he performs under the various forms of instructional method and in varying learning situations. It is this problem of matching students to learning experiences with which the present study is concerned.

The study was conducted as an adjunct to a research project sponsored by the Fund for the Advancement of Education.[2] The personality trait, sociability (or social introversion-extroversion), was studied in its relationship to academic achievement in an advanced educational psychology course. In establishing the various types of learning situations, an effort was made to devise situations representing various degrees of student interaction.

The general hypotheses regarding the relationship of sociability to course achievement which were investigated in this study may be stated as follows:

The more sociable student will achieve more under conditions permitting maximal student interaction (as in interactive discussion classes or in small group study) than will the less sociable student.

The less sociable student will achieve more under conditions of minimal student interaction (as in a lecture section or in totally independent study) than will the more sociable student.

Procedure and Subjects

The experimental *Ss* for the study consisted of the enrollees in the advanced educational psychology course, Psychology of Child Development, at the University of Michigan in the fall of 1956–57. The students enrolling in this course were a mixture of graduate and upperclass undergraduate students—a majority of them enrolled in the School of Education.

Following preliminary testing on course achievement and personality variables, the students were randomly distributed into the four experimental groups on a predetermined proportional basis (the reason for proportional distribution being to insure a smaller number of *Ss* in the discussion group and to make the lecture group the largest group). Group A, the lecture group, contained 36 students. They met regularly in traditional classroom fashion. Group B, the interactive discussion group, contained 13 students. They also met regularly with the instructor. Group C, the autonomous small-groups group, contained 25 students (later subdivided into small groups of five each). They met in small groups a minimum of nine times throughout the semester and did not have any contact with the instructor. Group D, the independent study group, contained 24 students. They studied independently

[1] Condensation of a paper presented at the joint meeting of the Oregon Psychological Association and the Washington State Psychological Association, May 9, 1959.

[2] Instructional Efficiency Project at the University of Michigan School of Education under the directorship of Warren A. Ketcham.

throughout the semester. The total experimental group numbered 98.

In order to further establish a baseline for comparison of achievement in the experimental groups, a control group was established. This group was comprised of 62 students in a sociology course (Marriage) in the College of Literature, Sciences and the Arts. They were given the same achievement tests as the experimental groups at the beginning and end of the semester. These students had not taken and were not currently taking the educational psychology course. Most of them were from the three upper classes of the Literature, Science, and Arts College, with a sprinkling of special and graduate students.

Measures Used

In the measurement of course achievement, a precourse test was given covering the content of the course. This test, composed by the author of the text, consisted of 60 multiple-choice items. An item analysis performed on these questions the preceding year indicated that they discriminated adequately among students taking the course. The identical test items were used as a postcourse test of achievement, but were scrambled in their order to overcome residual practice effects from previous experience with the test. The fact that several of the students in the course had had some work in education raised the lower limit of performance on the pretest but there was sufficient spread and ceiling on the posttest performance to confirm the test's effectiveness as an achievement measure. Achievement in the course was measured in terms of gain in scores on the posttest over the score earned on pretest.

The S scale (social introversion-extraversion) of Guilford's Inventory of Factors STDCR, used as the measure of sociability was selected because it is designed to assess the *action* dimension of the trait. Other scales such as those defining sociability as a "neurotic tendency" or "dependency need" were considered less suitable because the study was more concerned with actual propensity toward social interaction.

The Learning Situations

At the beginning of the semester all enrollees in the Psychology of Child Development course were tested on a number of variables pertinent to both the present study and the broader research project sponsored by the Fund for the Advancement of Education. All members of the group were assigned to the text, given an envelope of study material (including supplementary reading materials, a bibliography of readings, and a Study Guide for the course) and then assigned to their specific experimental groups.

Special care was taken to make the classroom sections typical lecture and interactive discussion sections, these two groups being handled by two different instructors. Both instructors find their respective methods of teaching the most "natural" for them. Each instructor is recognized by his colleagues as being skilled in employing the techniques peculiar to the teaching method he employed for the classroom section studied in this investigation.

To provide as adequate a picture as possible of what transpired in the various experimental groups during the course, a description of some typical activities and conditions within the groups is helpful. These descriptions are based on student reports and upon observations made on a time-sampling basis during the conduct of the class sessions.

Group A, The Lecture

The procedures used with the lecture group were generally quite structured. While the class was not a large lecture section, it was handled much the same as though it were. Occasional questions were raised by students but interaction among the students themselves was virtually nonexistent. The class met in a small auditorium and the

formal seating arrangement was conducive to instructor-centered proceedings. The lecturer displayed a somewhat informal manner but he did plan and direct activities, topics and presentations. The role of the student was primarily that of listener. This learning situation was validated as a traditional lecture situation through direct observation and evaluation on several criteria gleaned from a survey of pertinent research.

Group B, The Discussion

The classroom situation for Group B differed greatly from that of Group A. The class met in a small room with seats arranged about a large, low table. The physical setting was one generally conducive to relaxed informality. Coffee and tea were available during the class meeting. The general atmosphere and proceedings of the class meeting were definitely informal and there was a great deal of interaction among group members. The role of the instructor was principally that of one of the group. On occasion he did help to bring into the discussion some group members who were reticent in their participation. The role of the student in Group B was obviously one of active participant in the discussion. The group was small (14 including the instructor) and there was ample opportunity for opinions and ideas to be expressed and evaluated by all members of the group. Some special techniques such as role-playing, case studies, buzz groups, etc., were employed throughout the semester to enhance the value of the discussions and to encourage active student participation. This situation, also, was validated on the basis of several criteria found in the literature to be characteristic of an interactive classroom discussion group.

Group C, Autonomous Small Groups

It is more difficult to get a representative picture of the procedure followed in the last two experimental groups because they were out of contact with the instructor and established their own procedures for study. However, through *postmeeting reaction reports* which were turned in by each student after his group meeting and also through private conversations with group members, some information was obtained from the members of the autonomous study groups. Most of the small groups met more than the required minimum number of times. The meeting places were usually chosen on the basis of comfort or convenience. Typical meeting places were vacant and centrally located classrooms on the campus, the student union building, or the home or room of one of the group members. Many welcomed the opportunity to share more completely the different views and experiences represented within their small group and there was apparently very active participation by all members. As was intended, there was virtually no contact with the instructors.

Group D, Independent Study

Because of the rare contact with instructors or anyone connected with the study, little is known of the procedures followed by the students in Group D. Some did mention later that they spread their study over the whole semester while others found it more profitable to concentrate their study at certain points. Some reported that they "let the studying slide" until near the end of the semester.

Results

The precourse measures revealed that there were no significant differences from group to group on either the sociability scale or the achievement test. Checks on other factors such as intelligence test scores, previous courses in psychology and educational psychology, etc., revealed little difference in these areas.

Postcourse measures indicated that achievement gains in all experimental groups were significant beyond the .001 level of confidence. As shown in Table 1, the mean gain of each experimental group was significantly greater (also beyond the .001 level of confidence in each case) than the mean gain of the control group. A small (nonsignificant) gain found in the control group was considered attributable to the fact that some discussion on child development entered into the marriage course in the discussions on child rearing and family relations.

Turning to a comparison of the more sociable student with the less sociable student, an analysis is here presented in terms of the within-groups differences. For purposes of this analysis, the more sociable students in a given experimental group were those falling above the median of that

TABLE 1

ACHIEVEMENT GAIN IN EXPERIMENTAL GROUPS AND CONTROL GROUP

	Control Group	Group A (Lecture)	Group B (Discussion)	Group C (Small Groups)	Group D (Independent Study)
Mean Achievement Gain	2.7	10.72	8.31	8.84	9.75
Sigma	4.7	4.36	3.62	4.48	5.99
t (compared with control group)	—	8.44	4.68	5.63	5.07
t (net gain in achievement)	—	8.47	4.72	7.34	4.35

Note.—All t's are significant at better than the .001 level of confidence.

TABLE 2

SOCIABILITY AND ACHIEVEMENT IN THE EXPERIMENTAL GROUPS

	Group A (Lecture)				Group B (Class Discussion)		
	More Sociable	Less Sociable			More Sociable	Less Sociable	
High Achievers	6	12	18	High Achievers	1	5	6
Low Achievers	12	6	18	Low Achievers	5	1	6
$\chi^2 = 2.78$ ($p < .05$)	18	18	36	$\chi^2 = 3.00$ ($p < .05$)	6	6	12
	Group C (Small Groups)				Group D (Independent)		
	More Sociable	Less Sociable			More Sociable	Less Sociable	
High Achievers	10	2	12	High Achievers	6	6	12
Low Achievers	2	10	12	Low Achievers	6	6	12
$\chi^2 = 8.17$ ($p < .01$)	12	12	24	$\chi^2 = 0.17$ (ns)	12	12	24

group on the sociability measure and the less sociable students in a given experimental group were those falling below the median of that group on the sociability measure. This dichotomizing procedure was followed with achievement scores also. The members of each experimental group can then be tabulated in a two by two table in terms of relative sociability and relative achievement. The chi square test of significance (one-tailed), with correction for continuity, was applied for each group.

Table 2 shows the results of these tabulations in two by two tables for the four experimental groups.

When the chi square test was applied to these observed differences in achievement between the more sociable and less sociable students, significant differences in the predicted direction were found in Group A and Group C. In Group A, a difference in achievement was found (p less than .05) in favor of the less sociable student. In Group C, a very significant difference was found (p less than .01) favoring the more sociable student. Thus, in these two experimental groups the general hypotheses of the study find support.

The results in Group B (the classroom discussion group) are quite puzzling. The difference here is in the direction opposite to that predicted. Discussions with the instructor of this group provided few leads as to how the results might possibly be accounted for. The most promising hypothesis which might be worthy of further in-

vestigation is that the presence of the instructor plays a significant role in determining participation and the amount or the kind of participation by the more sociable student may vary with the instructor's presence or absence. The same may be true of the less sociable student. It is also possible that if other "course outcomes" had been evaluated in addition to achievement gain in course content, this type of result may not have appeared uniformly in Group B.

No differences in achievement were found between the more or less sociable students under conditions of completely independent study.

Summary

In studying the relationship of sociability to academic achievement in a course in advanced educational psychology, four experimental groups were established. These groups, characterized by varying degrees of student interaction and student-teacher contact, were: A (a lecture section), B (an interactive discussion class), C (a group of small instructorless groups of five members each), and D (an independent study group). In learning situations A and C a significant relationship was found between sociability and achievement as hypothesized, i.e., the less sociable student achieved more than the more sociable student in the lecture section and in the small-groups group the more sociable student achieved more than the less sociable student. It appears that in situations similar to the course here described the S (social introversion-extraversion) scale of the Guilford Inventory of Factors STDCR may have some value in identifying those individuals who will perform well in these two types of learning situations.

REFERENCE

Birney, R., & McKeachie, W. The teaching of psychology: A survey of research since 1942. *Psychol. Bull.*, 1955, **51**, 51-68.

(Received December 28, 1959)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 213-216

VARIATION OF WHOLE-PART METHODS OF LEARNING

F. J. McGUIGAN
Hollins College

McGuigan and MacCaslin (1955) in comparing the effectiveness of a part, an incomplete-whole and a whole method of learning found that the whole method was superior to the incomplete-whole method which in turn was superior to the part method. By holding the kind of practice (massed or distributed) of the culminating aspect of a task constant their comparisons led to the conclusion that the whole method is superior to the part method largely because the whole method involves distributed practice. An extension of their reasoning led to the hypothesis tested in this study: within limits, the larger the unit practiced (and hence the greater the distribution of practice on the various aspects of the task) the better the performance.

A secondary purpose was to study the effect of varying experimenters. Psychologists tacitly assume that their results may be generalized to a population of *E*s, but there has been relatively little sampling of such a population to justify the generalization.[1] For if interactions between experimenters and treatments exist, the extent to which experimental results may be generalized must be sharply restricted.

METHOD

Subjects. Seventy-two undergraduate women at Hollins College were used as *S*s. Nine *E*s who were women undergraduate majors in psychology were trained in the experimental procedure. The *S*s were randomly assigned to four groups and each *E* was randomly assigned to run two *S*s in each group.

Procedure. The task for the *S*s was to reproduce card Design Number XVII of the Kohs Block Design Test (C. H. Stoelting Co., Chicago, Illinois). The four groups were randomly assigned to four conditions: (*a*) part method, (*b*) very-incomplete-whole-method, (*c*) incomplete-whole-method, and (*d*) whole method. The card design was divided into four quadrants. The various quadrants were covered with pieces of black paper as appropriate for each condition. The procedures for the four groups were:

Group P (*Part Method*). *S*s for the part method were given four blocks and allowed to look at only the first quadrant of the design, the remaining three quadrants not being exposed to them. Then they reproduced the first quadrant with the blocks 12 times in succession. Following this the second quadrant was exposed while the other three quadrants were hidden from view. They then reproduced the second quadrant 12 times. In similar manner the third and fourth quadrants were reproduced 12 times apiece.

Group VIW (*Very-Incomplete-Whole-Method*). The *S*s in this group were given eight blocks and allowed to look at half of the design, vertically divided. They were instructed to reproduce the half of the design that they saw 12 times. Following this the first half of the design was covered and the second half exposed. The *S*s then reproduced the second half of the design 12 times.

Group IW (*Incomplete-Whole-Method*). The design was divided into two unequal parts for these *S*s. First the first, second, and third quadrants were simultaneously exposed while the fourth quadrant was covered. The *S*s reproduced these three quadrants as a unit 12 times. Then the first three quadrants were covered while the fourth quadrant was exposed. The *S*s then reproduced the fourth quadrant 12 times.

Group W (*Whole Method*). No part of the design was ever covered for the whole method *S*s. Hence, they reproduced the design 12 times as a total unit.

All *S*s were allowed to complete the design, or portions thereof, at their own pace.

[1] The necessity for sampling populations that are usually held constant in an experiment (such as a population of *E*s) has been pointed out by Hammond (1954) and Brunswik (1956).

Following completion of the above training all *Ss* were presented the 16 blocks and were allowed to look at the entire design. They were then instructed to reproduce the design as rapidly as possible. The time required for each *S* to complete the design on this test trial was recorded.

Using four methods and nine *Es*, it can be seen that a 4 × 9 factorial design was employed. The methods constitute a fixed variable while *Es* may be considered to be a random variable. Hence the analysis is that for a mixed model.

Results

The results of the analysis of variance are presented in Table 1 where it can be seen that the only significant *F* is for methods. The mean times on the test trial for Groups W, IW, VIW, and P were 65.85, 93.87, 167.87 and 166.53 seconds respectively. With the exception of the reversal for Group VIW and Group P, the order of means is that hypothesized. Appropriate testing indicated no basis for assuming that the variances were other than homogeneous. Duncan's Range Test (Duncan, 1955) was applied to determine significant differences among the various pairs of means. It was found that Groups W and IW were both superior to Groups VIW and P at the 5% level. No other differences were significant. It may thus be concluded that the methods that allowed the larger practice units (the whole method and the incomplete-whole-method) were superior to those that permitted practice on the smaller units (the very-incomplete-whole-method and the part method).

TABLE 1

Analysis of Variance for Time Scores on Test Trial

Source of Variation	*df*	*MS*	*F*
Between Methods	3	48,123.38	6.01*
Between Experimenters	8	2,539.12	.32
M × E	24	8,012.26	.93
Error	36	8,580.59	

* Significant beyond 1% level.

The lack of a significant interaction between methods and *Es* increases our confidence in generalizing to a population of *Es*. However, since the population of *Es* sampled was rather restricted, generalization to a larger or somewhat different population must be made with caution.

Discussion

The present status of research on the whole-part problem is well summarized by Underwood (1949):

> Results have been contradictory, sometimes showing facility in learning by wholes and sometimes by parts, and psychologists have been unable to specify adequately the nature of the differences between experiments which would account for discrepancies. The reason for this is probably that no comprehensive hypothesis concerning the variable has been set up by which the success of the studies may be evaluated ... there is a lack of adequate theory to direct experimentation (p. 418).

However, a review of the relevant literature indicates few results that would be accepted by present day scientific standards. For frequently *n*'s have been extremely small, appropriate tests of significance have not been applied, confounding of variables has been present, etc. The sounder research might lead us to conclude that the whole method is *generally* the superior method. The large number of variables that are relevant to the whole-part problem, however, would certainly leave room for exceptions (cf. Crafts, 1930). In an effort to bring order to the whole-part problem, as Underwood suggests is necessary for further research to be fruitful, the following theoretical basis is offered as at least a partial explanation for the apparent superiority of the whole method.

The results of the present investigation are consonant with the previously advanced hypothesis that much of the superiority of the whole method is due to the fact that it entails distributed practice. To

illustrate this conclusion by means of the present task, let us say that R_1 is the response of placing the blocks together to form the first quadrant of the design, R_2 is the response for constructing the second quadrant, etc. Now, when the part method is used, R_1 is made repeatedly, then R_2 is made repeatedly, etc. Hence, practice on R_1, R_2, etc. may be said to be massed. The whole method, on the other hand, entails distributed practice on the four parts of the task. Thus, R_1 is made once followed by R_2, then R_3 and finally R_4, or whatever order S selects. It can thus be seen, from this point of view, that the whole-part problem becomes one concerned with massing vs. distribution of practice. And Hull (1951) has offered us a reasonable explanation of why distributed practice is superior to massed practice. Applying his theory to the whole-part problem, using the present task as an example, we may say that when the part method is used, the reactive inhibition entailed by continuous performance of R_1 is sizeable (Hull's Postulate IXA). However, when using the whole method, R_1 is followed immediately by R_2. Therefore, the reactive inhibition accumulated by performing R_1 with the whole method is rather small; in particular it is smaller than when R_1 occurs repeatedly as with the part method. And so on for the remaining responses. Following Hull, the greater the amount of reactive inhibition present at any given time, the greater the amount of conditioned inhibition (Hull's Corollary IX). Since with the part method, practice on every response is massed (R_1 for instance), the amount of reactive inhibition present at the end of training is greater than for the whole method (where with distributed practice on R_1 for instance, reactive inhibition dissipates to a greater extent). Therefore, the amount of conditioned inhibition attached to any response would be greater when the part method is used than for the whole method. This means at any given time (either immediately after the task has been finished, or after considerable time when reactive inhibition has dissipated) the inhibitory potential for R_1, R_2, R_3, and R_4 will be greater when the part method has been used than when the whole method has been used. For this reason, other factors being equal, we should expect the whole method to be superior to the part method. Of course this extension of Hull's theory must be checked by further test, and since other factors are not always equal, we should expect some exceptions. Further research should enable us to better specify the effect of these other relevant variables.

The above reasoning was illustrated by using the extreme methods employed in the present study, the whole and part methods. A general application of the extension of Hull's theory to all groups would be as follows. First let us assume the degree of distribution of practice for the four groups in increasing order is: Group P, Group VIW, Group IW, and Group W. That is, Group P had the most massed practice on each response while Group W had the most distributed practice on each response. Therefore Group P should have generated the largest inhibitory potential for each response, Group VIW the next largest, Group IW the next, and Group W the least. Therefore, it should be expected that Group P would perform the poorest, Group VIW the next poorest, Group IW the next, and Group W should exhibit the best performance. The results were in general agreement with this expectation.[2]

It should be pointed out that the above theorizing rests on the assumption that the inhibitory potential generated by a given

[2] As with any known experiment, alternative explanations of the present data may be expected. That a theory postulating that the amount of "integration" of the design is not appropriate is suggested by work on this variable by Seagoe (1936). For where she varied the "integration" of the design reproduced the only difference between whole and part methods that even approached significance were where the whole method was superior.

response is uniquely attached to that response. That is, except for possible generalization of inhibitory potential among responses, the amount of inhibitory potential relevant to each response is that generated by it and by no other response. This assumption is crucial to the above explanation because if the reactive inhibition generated by R_1 is also attached to R_2, R_3, and R_4 then continuous practice by the whole method would result in the same inhibitory potential as by the part method.

Summary

This study sought to test the hypothesis that the larger the unit practiced, the better the performance; and also to study the effect of varying experimenters. Four methods which varied the extent to which part or whole learning was possible were used. Nine *Es* ran 72 *Ss*, a total of 8 *Ss* each. The results for methods showed, in general, that the larger the practice unit, the better the performance and were thus in accord with the previously advanced hypothesis that much of the superiority of the whole method may be due to the fact that it entails distributed practice. Hull's theory of inhibition was extended to provide a theoretical basis for explaining the superiority of the whole method over the part method. The results also showed that variation of *Es* did not lead to an interaction with methods. This finding increases our confidence in generalizing results for methods to a hypothetical population of experimenters, though more thorough sampling of such a population should occur in future research.

REFERENCES

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley & Los Angeles: Univer. Calif. Press, 1956.

Crafts, L. W. Whole and part methods with non-serial reactions. *Amer. J. Psychol.*, 1930, **41**, 543–563.

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

Hammond, K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, **51**, 150–159.

Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

McGuigan, F. J., & MacCaslin, E. F. Whole and part methods in learning a perceptual motor skill, *Amer. J. Psychol.*, 1955, **68**, 658–661.

Seagoe, M. V. The influence of degree of wholeness on whole-part learning. *J. exp. Psychol.*, 1936, **19**, 763–768.

Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

(Received January 14, 1960)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 217-221

THE EFFECTS OF VISUAL DISCRIMINATION PRETRAINING ON LEARNING TO READ A VOCABULARY LIST IN KINDERGARTEN CHILDREN[1]

SIEGMAR MUEHL

Iowa Child Welfare Research Station

To master beginning reading, a child must be able to associate word names, or meanings, with their printed symbols. To make these associations a beginning reader must be able to see and respond to the printed symbols themselves as distinctive and stable stimulus compounds.

How does this visual discrimination skill develop? Theories of perceptual development attempt to relate sensory input to the learning process. The perceptual theories developed by such writers as Hebb (1940), Gibson (1940), and Murphy (1951) stress the contribution of sensory experience to perceptual differentiation of a stimulus field. Thus, for beginning readers, the question can be asked: What types of sensory experiences facilitate visual discrimination of complex stimuli represented by printed words in a reading task? The answer has practical implication for visual discrimination training in the reading readiness program. McKee (1948, p. 147) has noted that commercial reading readiness books concentrate on visual discrimination exercises using geometric or animal forms rather than printed words. Educational research (Goins, 1958) has never confirmed the effectiveness of this type of discrimination training for beginning reading.

Although not directly concerned with the learning process defined in terms of reading performance, psychological studies in the area of discrimination pretraining provide hypotheses, research procedures, and findings which may contribute to understanding the relation between visual discrimination skills and reading performance. Studies in this area have investigated the effects of different types of stimulus discrimination pretraining on performance in a subsquent discrimination learning task. These studies provided tests of either Miller's (1948) acquired distinctiveness of cues hypothesis, or of Gibson's (1940) transfer of discrimination hypothesis. Both hypotheses state, in effect, that if differences are established among similar stimulus items or situations, then learning will be facilitated in a subsequent task using these same stimuli. An analysis of this research by Spiker (1956), and the later findings of Kurtz (1955), indicated that stimulus differentiation pretraining does result in positive transfer to a second task employing the same visual stimuli but different responses as used in the pretraining task.

The general objective of the present study was to extend the experimental procedures used in the stimulus pretraining research to investigate the question raised above: What types of sensory experiences facilitate visual discrimination among printed words occurring in a reading task? Specifically, the purpose was to investigate the effects of different types of visual discrimination pretraining on the performance of kindergarten children in learning to read a vocabulary list. Prior to learning the same vocabulary list, kindergarten children were given matching pretraining with three types of stimulus material: the same words as those appearing in the vocabulary list; different words; and geometric forms. Based on the results of the

[1] The author wishes to thank Buford W. Garner, Superintendent of Schools in Iowa City, John Moore and Richard Hovet, Principals, and Lois McDermott and Nancy Cooker, kindergarten teachers, whose cooperation made this research possible.

previous experimental work in stimulus pretraining, the prediction was made that the group given visual discrimination pretraining matching the same words would learn the vocabulary list faster than the groups given pretraining with different words or geometric forms.

METHOD

Pretraining

Three different sets of stimuli were used in the matching pretraining, with four stimuli in each of the sets. There were two random orders for each set of four stimuli. Set S (same words) consisted of four words: BOAT, HELP, PLAY, COME. Set D (different words) constisted of four words: MAKE, JUMP, WORK, FIND. Set F (control) consisted of four geometric forms. Differences in these forms were obtained by varying a pentagonal figure with a single line connecting different inside corners. Set S and D stimulus words were typed in lower case with a primary style typewriter. Set F stimulus forms were drawn with ink and measured 1¼″ on the base and in height. The individual stimuli in all sets appeared on a 2¼″ × 2¾″ white card. Accompanying each order of stimuli in each set were two, 2¾″ × 8½″, multiple-choice response cards. The four stimuli appearing on the stimulus cards of the set also appeared in two different arrangements on the two response cards.

The apparatus for presenting the stimulus and response cards consisted of a sloping metal surface, 8½″ × 9″. Close to the top of this surface and approximately at the eyelevel of the *S* was a small ledge for presenting the stimulus cards. The bottom three inches of the metal surface consisted of a hinged window which could be raised by the *S*. A slot behind the window held a response card. The latter could be placed in position from behind without the *Ss* seeing it. Directly in front of the response card window was a flat, 3″ × 9″ panel containing four buttons. When the response card window was open, these buttons were directly in front of each of the four stimuli on the response card. The buttons were placed in a circuit with a selector switch and chime located behind the apparatus. The switch enabled the *E* to select any one of the four response buttons. When the correct button was pressed by *S*, the chime sounded.

Three groups of *Ss* were differentiated with respect to their pretraining experience. *Ss* in Group S were given six trials with the four Set S stimuli. *Ss* in Group D were given six trials with the four Set D stimuli. *Ss* in Group F were given six trials with the four Set F stimuli. A presentation for all groups consisted of *E*'s placing a stimulus card in the card holder for approximately two sec. *S* was told to look carefully at the card. At the end of this interval, *E* removed the stimulus card and *S* was instructed to open the response card window, find the same stimulus on the response card that had appeared on the stimulus card, and press the button directly under this response stimulus. The *S* had approximately five sec. in which to press the button. If *S* pushed the correct button the chime would ring. If *S* pushed the wrong button, or if *S* failed to respond in the approximately five-sec. interval, *E* closed the response card window and a new stimulus card was presented. *E* changed the response card after each order of stimulus cards had been presented. No rewards were given during pretraining other than verbal encouragement from *E* and the reinforcement provided by the chime.

Learning Task

Immediately following the pretraining matching task, the vocabulary learning task was administered. All *Ss* received 12 trials with the four stimuli of the Group S set. The *Ss* were instructed that they were to try to learn what the words said. On the first trial *E* presented each of the four stimuli and told *S* what each said. On the following trials, *S* was instructed to

try to beat *E* in saying the word. Following the first trial *E* used a paired-associate method of presenting the stimulus and response elements. *E* presented the stimulus card. *S* was given an approximately two-sec. anticipation interval in which to say the word. At the end of this interval *E* gave the correct response, with the stimulus card still present to the *S*. No rewards were given during the learning task other than verbal encouragement from *E* and the reinforcement provided by *S*'s making a correct anticipation. Upon completion of the learning task, *E* asked the *Ss* in Groups S and D to tell how they matched the words in the pretraining task.

Subjects

The *Ss* were 37 children from three kindergartens in the public schools in Iowa City. *Ss* were randomly assigned to the three pretraining groups. There were 13 *Ss* in Group S; 12 *Ss* in Group D; and 12 *Ss* in Group F. Ages ranged from 5–1 to 6–2. The mean age was 5–6. There were no significant age differences among the three pretraining groups. Testing was done in December.

Results

Pretraining

The means and standard deviations of correct responses in the six pretraining trials are given in Table 1. The mean number of correct responses for the three pretraining groups did not differ significantly among the groups. (Lindquist [1953] Type I analysis, $p > .05$). The mean number of correct responses over trials for the combined groups did differ significantly ($p < .01$). An analysis of the correct response trend over trials indicated that discrimination learning occurred during pretraining.

TABLE 1

Number of Correct Responses in Six Pretraining Trials

	Group S	Group D	Group F
Mean	18.23	17.75	17.08
SD	3.55	3.70	1.93

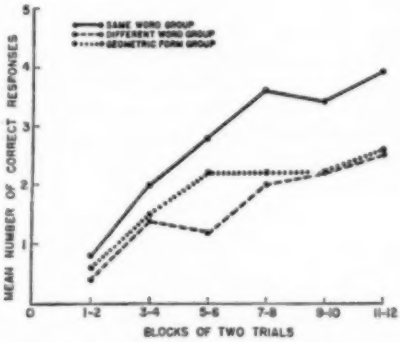

Fig. 1. The effect of type of matching pretraining on correct responses in learning the four word vocabulary list. Eight responses were possible in each block of two trials.

TABLE 2

Analysis of Variance of Correct Responses in 12 Learning Trials

(in Blocks of Two Trials)

Source	*df*	*MS*	*F*	*p*
Between-Subjects	36			
Groups (B)	2	25.88	3.74	$<.05$
Error (b)	34	6.91		
Within-Subjects	185			
Trials (A)	5	28.66	24.07	$<.01$
A × B	10	1.41	1.18	*ns*
Error (w)	170	1.19		
Total	221			

Learning Task

The learning curves in the vocabulary task for each of the pretraining groups are presented in Fig. 1. A summary of the Lindquist (1953) Type I analysis of variance for these data is given in Table 2. As shown in the summary, type of matching pretraining significantly affected learning performance ($p < .05$). The interaction

of type of pretraining with trials was not significant ($p > .05$). When Groups S and D, and S and F were compared in learning performance the differences were significant ($p < .05$). Comparing Groups D and F, the difference in learning performance was not significant ($p > .05$).

Discussion

The prediction was made that *Ss* given visual discrimination pretraining matching the same words to appear in a vocabulary list would learn this faster than *Ss* given matching pretraining with either different words or geometric forms. The results supported this prediction. Since there was no no-discrimination pretraining control group, it was not possible to determine whether positive transfer resulted from the different word and geometric form pretraining.

Ss' reports on how they matched the words, and an analysis of the type of intraset errors made during pretraining, provided some evidence to account for these results. Single letters or details of these letters appeared to be the basis for word matching. The characteristics that made a letter stand out visually were position within the word, most often first and last letters, and unique shape. This evidence suggests that the positive transfer to the learning task in Group S was due to the acquisition, during pretraining, of a set of observing responses to distinctive letters in the stimulus words. These learned responses served to make the words in the vocabulary list more distinctive at the beginning of learning for Group S than for Groups D and F. It is not possible to characterize the precise nature of these observing responses from the present data. With verbal *Ss* it is likely that these responses are mediated by implicit verbal responses. In support of this latter possibility, a few *Ss* reported they matched the words by looking at first letters, or by a letter which was in their name. There was no indication, however, that the majority of *Ss* knew, or were using letter names as a basis for mediating the matching process.

In comparing Group D with Group F, the evidence showed that the acquisition of a set of observing responses suitable to differentiate one list of words did not significantly transfer, either negatively or positively, to the task of acquiring another set of observing responses with a different list of words.

Summary

This experiment compared the performance of three groups of kindergarten children in learning to read a vocabulary list. The groups differed in their pretraining experience. Group S learned to match the same words that appeared in the vocabulary list; Group D matched different words; and Group F matched geometric forms.

The response measure was the number of correct anticipations in 12 trials of the vocabulary learning list. Analysis of the data showed that Group S performed significantly better than Groups D and F. The difference between Groups D and F was not significant.

The results suggest that word recognition skill is related to the development of specific visual discrimination skills. The results also agreed with the findings of previous psychological research in stimulus differentiation pretraining.

REFERENCES

Gibson, Eleanor J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

Gibson, J. J., & Gibson, Eleanor J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–41.

Goins, Jean T. Visual perceptual abilities and early reading progress. *Suppl. educ. Monogr.*, 1958, No. 87.

Hebb, D. O. *The organiaztion of behavior: A neuropsychological theory*. New York: Wiley, 1940.

Kurtz, K. H. Discrimination of complex

stimuli: The relationship of training and test stimuli in the transfer of discrimination. *J. exp. Psychol.*, 1955, **50**, 283–292.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCKEE, P. *The teaching of reading in the elementary school.* Boston: Houghton Mifflin, 1948.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

MURPHY, G., & HOCHBERG, J. Perceptual development: Some tentative hypotheses. *Psychol. Rev.*, 1951, **58**, 332–349.

SPIKER, C. C. Experiments with children on the hypothesis of acquired distinctiveness and equivalence of cues. *Child Develpm.*, 1956, **27**, 253–261.

(Received January 25, 1960)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 222-234

COLLEGE PRESS AND CHANGES IN STUDY PLANS OF TALENTED STUDENTS

DONALD L. THISTLETHWAITE
National Merit Scholarship Corporation

Despite its importance, the concept of "recruiting" able persons for the intellectual disciplines will seem alien to some readers. The notion of proselytising violates their self-concepts. At best, persuasion seems an ineffective means of attracting promising young scholars and scientists into higher educational programs. A broader use of "recruitment" is considered here: more generally, the process of attracting talented undergraduates into different careers may be seen to consist—not in persuading them to do things they do not wish to do, but rather—in helping them to discover how they can best realize their creative and productive potentialities.

In the present investigation, changes in major field of study and in the level of training sought, were related to the college experiences of winners and near-winners in a national scholarship competition. Information was obtained on the relative "holding power" and attractiveness of different fields of study, the characteristics of students and faculty in these fields, the effects of intellectual and social atmospheres upon motivation to seek advanced training, and the traits of instructors which had the greatest influence upon these talented students. An understanding of the social systems influencing students to specialize in certain fields and to seek graduate or professional training might enable us to create college environments in which more students could develop their talents.[1]

[1] This research was partially supported by the National Science Foundation, the Old Dominion Foundation, and the National Merit Scholarship Corporation. The author is indebted to Laura Kent and to John L. Holland for a critical reading of the earlier version.

Procedure

Sample. In May, 1959, 546 winners (Merit Scholars) and 1,480 near-winners (finalists who received Certificates of Merit) in the 1956 National Merit Scholarship competition were surveyed by mail. The survey was limited to students who had responded to previous surveys and who were enrolled in college. Of those to whom questionnaires were sent, 117 were considered ineligible for inclusion in the study. (This group consisted of 40 who returned incomplete questionnaires, 44 who had withdrawn from college, 12 who had transferred to a new college, 17 who had moved and left no forwarding address, 3 who had entered religious seminaries, and one who died.) Returns were received from 1500 of the eligible students: 469 Scholars (92% returns), and 1,031 C of M winners (75% returns).

All of the 1500 respondents took the College Entrance Examination Board tests during their senior year in high school (1956); 99% had SAT verbal scores and 84% had SAT mathematical scores which placed them above the 95th percentile relative to the nation's high school seniors (Levine, 1958). Two-thirds of the group had verbal scores and half had mathematical scores which placed them above the 95th percentile relative to typical College Board examinees. The median SAT verbal score for the group was 685, and the median SAT mathematical score was 691.

The sample—987 men and 513 women—includes students from 327 different American colleges and universities.

Changes in study plans. Information about changes in the direction of study plans was elicited using a list of 61 sub-

ject-fields, identical to those used by Iffert (1957). The student was instructed to:

> Write the number (1) before the subject in the list below which corresponds to your *present major field of study*. If your major interest has changed since you entered college write the number (2) before the subject in which you planned to major when you entered college.

To estimate the impact of undergraduate experiences upon plans to seek advanced training each student was asked, "Has the level of training you aspire to achieve changed since you entered college?" The three fixed alternative responses to this question are shown in Table 3.

Classification of fields of study. The student's field of concentration was classified into one of five broad areas: natural sciences (astronomy, chemistry, engineering geology, mathematics, metallurgy, physics); biological sciences (agriculture, bacteriology, biochemistry, biology, botany, nursing, physiology, predentistry, premedicine, zoology); social sciences (anthropology, economics, education, geography, history, international relations, political science, psychology, social work, sociology); arts and humanities (architecture, English, fine arts, language, music, philosophy, speech, dramatic arts); and unclassified fields (accounting, business administration, home economics, journalism, prelaw, religious education, and other specialties not elsewhere classified). Within each of these broad areas the most popular fields of study—at the end of the third year of college—were as follows: engineering (35% of the students in natural sciences); biology (24% of the students in biological sciences); history (35% of the students in social sciences); and English (49% of the students in the arts and humanities).

Faculty and student press scales. Thistlethwaite (1959a, 1959b) has shown that variations between colleges in the proportions of their alumni who subsequently earned the PhD are related to differences in intellectual and social atmospheres. In the earlier study, selected items from the College Characteristics Index (Pace & Stern, 1958; Stern, 1958) had been grouped on the basis of similarity of content, taking care to make each item-cluster descriptive of either faculty or students but not both; items were then discarded which (*a*) did not differentiate between "high" and "low" productivity colleges, or (*b*) had low correlations with total provisional scores of items with which they were clustered on the basis of content. For the purposes of the present study, additional items were written for each cluster and each item-cluster edited so that each scale contained 10 items, five keyed so that a True response is weighted positively and five keyed so that a False response is weighted positively. Scale reliabilities (Kuder-Richardson Formula 20), sample items, and responses weighted positively for each of the items are shown in Table 1.[2] In answering faculty press items, students were instructed to describe the facilities, faculty, and requirements in their *present major field of study*; in answering student press items, they were instructed to describe the activities and interests of students in their *living group and classrooms*.

Results

Changes in Study Plans

Major fields of study. During the three years 1956–59 the proportions of boys and girls majoring in the natural and biological sciences decreased while the proportions majoring in the social sciences and humanities increased (Table 2). Although the

[2] Complete lists of items in each scale, and keys used in scoring items have been deposited with the American Documentation Institute. Order Document No. 6328, remitting in advance $1.75 for microfilm or $2.50 for photocopies. I am indebted to C. Robert Pace and George G. Stern for permission to incorporate items from their College Characteristics Index in these scales.

TABLE 1

RELIABILITIES AND REPRESENTATIVE ITEMS OF FACULTY AND STUDENT PRESS SCALES

Press Scale	Most Representative Item (Highest item-total score correlation)	Scale Reliability
Faculty scales:		
Achievement	Standards set by the professors are not particularly hard to achieve. (F)	.73
Affiliation	The professors really talk *with* the students, not just at them. (T)	.70
Compliance	Frequent tests are given in most courses. (T)	.57
Directiveness	Instructors clearly explain the goals and purposes of their courses. (T)	.50
Enthusiasm	Lectures and class discussions are typically vigorous and intense. (T)	.77
Humanism	Student interest in understanding and criticizing important works in art, music, and drama is encouraged by the faculty. (T)	.75
Independence	There is little opportunity for pursuing independent study under the supervision of faculty members. (F)	.60
Pragmatism	The academic atmosphere is practical, emphasizing efficiency and usefulness. (T)	.69
Supportiveness	Some professors here tend to belittle the students. (F)	.44
Student scales:		
Achievement	Most students here are genuinely interested in learning and studying. (T)	.74
Aggression	Personal rivalries are fairly common. (T)	.59
Breadth of interests	Most students have very little interest in round tables, panel meetings, or other formal discussions. (F)	.57
Competition	The competition for high achievement is intense. (T)	.77
Humanism	There is a lot of interest here in poetry, music, painting, sculpture, architecture, etc. (T)	.78
Participation	Student pep rallies, parades, dances, carnivals or demonstrations occur very rarely. (F)	.57
Reflectiveness	Books dealing with psychological problems or personal values are widely read and discussed. (T)	.76
Scientism	When students get together they seldom talk about science. (F)	.76
Social Conformity	Proper social forms and manners are important here. (T)	.62

retention rate for boys in the natural sciences (80.6%) was significantly greater ($p < .01$) than for boys in the biological sciences (64.1%) or in the arts and humanities (68.6%), male enrollment in the former fields decreased. The explanation for this paradox lies in the failure of the natural sciences to recruit new students. The natural sciences and the biological sciences recruited a significantly smaller proportion ($p < .01$) of their present male majors from other fields than did the social sciences or arts and humanities. For example, the social sciences recruited 3.12 new males for each boy who withdrew, while the natural sciences recruited only one new male student for each three who withdrew. The retention rate for girls in the arts and humanities (75.2%) was significantly greater ($p < .01$) than for girls in the natural sciences (60.0%), social sciences (59.8%), or biological sci-

TABLE 2

DISTRIBUTIONS OF STUDY PREFERENCES OF 987 BOYS AND 513 GIRLS AT BEGINNING OF COLLEGE AND THREE YEARS LATER

Field of Study	Percentage Majoring in Each Field in		Percentage of Initial Majors Remaining in Field	Percentage of present Majors in Field Recruited from Other Fields	Ratio of Recruits to Withdrawals in Each Field
	1956	1959	(3 years later)		
Males (N = 987)					
Natural Sciences	62.8	54.6	80.6	7.2	.32:1
Biological Sciences	10.4	8.3	64.1	19.5	.43:1
Social Sciences	10.8	16.3	75.5	50.3	3.12:1
Arts and Humanities	8.7	11.8	68.6	49.1	2.11:1
Unclassified fields	7.3	9.0	59.7	51.7	1.59:1
Total (All fields)	100.0	100.0	75.8	24.2	
Females (N = 513)					
Natural Sciences	26.3	20.5	60.0	22.9	.44:1
Biological Sciences	12.3	11.1	50.8	43.9	.81:1
Social Sciences	19.9	22.2	59.8	46.5	1.23:1
Arts and Humanities	27.5	35.5	75.2	41.8	2.17:1
Unclassified fields	14.0	10.7	50.0	34.6	.51:1
Total (All fields)	100.0	100.0	61.6	38.4	

ences (50.8%). Like boys, girls found the natural sciences least attractive as a new field of study: the natural sciences recruited a significantly smaller ($p < .01$) proportion of their present female majors from other fields than did the biological sciences, social sciences or arts and humanities.[8]

Plans to enter graduate or professional schools. After three years of college, approximately one-third of these students are planning to do the three or more years of graduate study required for a PhD degree, while one in seven expect to get a professional degree in medicine or law. The proportion of students planning to get advanced degrees varies considerably from one field of study to another. For example, 73% of those majoring in the biological sciences expect to seek the PhD, MD, or similar degree; while 56% of natural science majors, 43% of the social science majors, and 36% of those majoring in arts and humanities aim this high.

Some of these differences undoubtedly can be attributed to factors other than college experiences. Several previous studies suggest that fields vary in the proportions of students who make their career-decisions before entering college (Gropper & Fitzpatrick, 1959; Sibley, 1948; Thielens, 1957). To remove the effects of such early decisions from comparisons of the proportions in each field making decisions to seek more advanced training, we have excluded all students who reported that they planned to seek one of the highest degrees (PhD, MD, etc.) and that their plans had *not* changed since entering college. These students pre-

[8] Comparisons of the percentages of boys maintaining their interests in specific fields over three years with those reported by Iffert (1957, p. 56) indicate that Merit finalists tend to be much more stable in their study plans than the typical college student. Iffert reports that the percentages of boys maintaining their interests in Chemistry, Engineering, Mathematics, and Physics are 21, 61, 20, and 29 respectively. For boys in the present sample the corresponding percentages are 66, 94, 39, and 73.

TABLE 3

REPORTED CHANGES IN HIGHEST LEVEL OF EDUCATION STUDENTS PLAN TO COMPLETE

Response to: Has the level of training you aspire to achieve changed since you entered college?	Percentage Giving Each Response among Students Majoring in Each Field[a]					
	Natural Sciences (N = 354)	Biological Sciences (N = 51)	Social Sciences (N = 184)	Arts and Humanities (N = 212)	Others (N = 101)	Total (N = 902)
No.	42	29	41	43	51	42
Yes. I now plan to seek a more advanced degree	47	65	48	49	43	48
Yes. I now plan to seek a less advanced degree	11	6	11	8	6	10
Totals	100	100	100	100	100	100

[a] Students who, in 1959, planned to seek the PhD, MD, or similar degree and who reported they had not changed their plans since entering college are not included.

sumably had made their decisions before entering college and their experiences could not result in their aspiring to a higher degree.

The remaining students, of which there were 902, could either (*a*) raise their level of aspiration or (*b*) seek a less advanced degree. Changes in aspirations, classified by major field of study, are shown in Table 3 for boys and girls combined. Approximately half of them report that they now plan to seek more advanced degrees.[4] Sixty-five per cent of those majoring in the biological sciences have raised their goals, while somewhat less than 50% of the students in other fields report such changes in plans; this difference is statistically significant ($p < .05$).

Before turning to an analysis of college experiences related to changes in the direction and level of career plans, we shall consider some of the differences between fields with respect to faculty and student press. Information about the intellectual and social atmospheres in difficult fields may help in understanding why many talented students change their study plans.

Characteristics of Students and Faculty in Specialized Fields

Student reports of the demands and pressures of the faculty in their major field and of students in their classroom and living groups indicate striking differences between the various specialized fields. Press scales have been grouped in Table 4 in clusters according to similarity of the pattern of deviations of the mean score for each field from the grand mean for all fields. It can be seen that 16 of the 18 press scales significantly differentiated between fields. In general, the natural sciences and the arts and humanities are most dissimilar in faculty and student press.

Natural sciences. The faculty in these fields is characterized by strong press for pragmatism, achievement, and compliance, by relatively little enthusiasm, affiliation, and supportiveness, and by weak press for humanism and independence. Student cultures of majors in these fields show strong press for scientism, competition and aggressiveness, and weak press for humanism, reflectiveness, breadth of interests, and social conformity.

Arts and humanities. Faculty and stu-

[4] Responses of the entire group of 1500 students indicate that 62% have not changed, 31% have raised, and 6% have lowered their aspirational levels. (One per cent gave no response to this item.)

TABLE 4

FACULTY AND STUDENT PRESS IN FOUR FIELDS OF STUDY

Press Scale	Deviation of Mean Score of Given Group from Grand Mean for Each Scale[a]				*F* ratio
	Natural Sciences	Biological Sciences	Social Sciences	Arts and Humanities	
Faculty scales:					
Humanism	*−166*	−33	+96	*+102*	154.2*
Independence	*−74*	−8	*+49*	+33	33.6*
Enthusiasm	*−71*	−11	+27	*+54*	21.3*
Affiliation	*−40*	−10	−6	*+54*	11.5*
Supportiveness	*−29*	+6	+9	*+13*	5.4*
Pragmatism	*+87*	+29	−36	*−82*	43.9*
Achievement	*+27*	*+27*	−19	*−36*	5.4*
Compliance	*+27*	+19	*−51*	+4	9.5*
Directiveness	−9	+26	−3	−14	1.7
Student scales:					
Humanism	*−67*	−38	*+58*	+47	21.3*
Reflectiveness	*−54*	−13	*+56*	+10	12.3*
Breadth of interests	*−38*	+10	*+24*	+4	7.9*
Social conformity	*−46*	+4	+4	*+36*	10.7*
Scientism	*+107*	+15	−39	*−85*	55.2*
Competition	*+37*	+1	+16	*−55*	9.0*
Aggressiveness	*+38*	+8	−17	*−28*	8.3*
Participation	+21	+1	*−45*	*+24*	7.3*
Achievement	−2	−8	+24	−15	1.5

[a] Grand means calculated weighting each field mean equally. Decimal points in deviation scores have been omitted (e.g., .27 is written 27). The highest and lowest deviation on each scale differentiating between fields is printed in italics.

* $p < .01$. All significance tests were made by two-tailed tests. The number of student observers in each broad field, reading from left to right, was 642, 130, 275, and 297, respectively.

dent press for majors in these fields generally present the most extreme contrast with the above traits. Teachers tend to be enthusiastic, affilitative, and supportive, to exert strong press for humanism and independence, and relatively weak press for pragmatism and achievement. Student norms and values are characterized by humanism, social conformity and participation, and by the absence of scientism, competition, and aggressiveness.

Biological sciences. Both teachers and students resemble those in the natural sciences though they do not differ as markedly from teachers and students in other fields as do those in the natural sciences.

Social sciences. Although the faculty and students in these fields resemble those in the arts and humanities on many scales, there are a number of exceptions. Specifically, social science faculties are characterized by weak press for compliance, and by relatively little affiliation, while the student cultures of social science majors emphasize reflectiveness, breadth of interests, and competition, placing little stress upon participation. Like the arts and humanities, the social sciences are associated with faculties which score high

on humanism, independence and enthusiasm, but low on pragmatism and achievement, and with student cultures relatively high on humanism but low on scientism.

College Press and Plans to Seek Advanced Training

Faculty behavior and press. Predictions made on the basis of the previous study (Thistlethwaite, 1959a, 1959b) are shown in the first column of Table 5. It can be seen that all 11 predictions were confirmed in that the groups predicted to have the higher means did in fact. Six of the 11 differences for which predictions were made were statistically significant, as was one difference which was not predicted. Thus, the following relationships have been confirmed in two separate studies: *college environments characterized by faculty affiliation or enthusiasm or faculty emphasis upon achievement, humanism, or independence are associated with increased motivation to seek advanced degrees in the arts, humanities, and social sciences; college environments characterized by a lack of faculty emphasis upon student compliance are associated with increased motivation to seek advanced degrees in the natural and biological sciences.* In addition, the present results indicate that teachers in the natural and biological sciences whom students perceive to be enthusiastic tend to be more effective in stimulating students to seek more advanced training.

Responses to the direct question, "What college experiences have influenced the level of training you plan to reach (e.g., BA, MA, PhD level)?" likewise indicated that college instructors frequently influence the student's decision to seek advanced training. Table 6 shows that 54% of the students who raised their level of aspiration mentioned "a few instructors who inspired me to emulate them," compared with only 13% of those who lowered their level of aspiration and

TABLE 5

PREDICTED AND OBTAINED DIFFERENCES IN FACULTY PRESS AS REPORTED BY STUDENT SAMPLES WITH DIFFERENT CHANGES IN LEVEL OF ASPIRATION

Faculty Press Scale[a]	Present Major Field					
	Natural Sciences and Biological Sciences			Arts, Humanities, and Social Sciences		
	Mean for *Ss* Increasing Level of Aspiration (N = 201)	Mean for *Ss* Not Changing Level (N = 162)	Mean Diff.	Mean for *Ss* Increasing Level of Aspiration (N = 191)	Mean for *Ss* not Changing Level (N = 166)	Mean Diff.
Achievement (+, +)	6.19	5.89	+.30	5.94	5.43	+.50*
Affiliation (+, +)	5.88	5.61	+.27	6.68	5.99	+.68***
Compliance (−, ...)	4.71	5.28	−.57***	4.20	4.21	−.01
Directiveness (−, ...)	7.24	7.39	−.15	7.27	7.07	+.20
Enthusiasm (..., +)	6.82	6.14	+.68**	7.87	7.32	+.55*
Humanism (..., +)	5.36	5.22	+.14	8.04	7.48	+.56***
Independence (..., +)	6.25	6.10	+.15	7.72	7.07	+.66***
Pragmatism (..., −)	5.93	6.16	−.23	4.25	4.48	−.23
Supportiveness (..., +)	5.46	5.21	+.25	6.08	5.68	+.40

[a] The first member of each pair of symbols in parentheses refers to the predicted direction of the difference among students majoring in the natural or biological sciences; the second member refers to the predicted direction of the difference among students majoring in the arts, humanities or social sciences. A dotted line (...) is shown where no prediction was made.

* $p < .05$ (one-tailed test).

** $p < .05$ (two-tailed test).

*** $p < .01$ (one-tailed test).

TABLE 6

COLLEGE EXPERIENCES AND THE LEVEL OF TRAINING STUDENT PLANS TO COMPLETE

Response to: What college experiences have influenced the level of training you plan to reach (e.g., BA, MA, PhD level)?	Percentage of Group Giving Each Response[a]			Difference between Columns 1 and 3
	Ss Who Decreased Level of Aspiration (N = 82) (1)	*Ss* Who Did Not Change Level of Aspiration (N = 365) (2)	*Ss* Who Increased Level of Aspiration (N = 427) (3)	
A few instructors who inspired me to emulate them	13.4	21.9	54.1	+40.7**
Evaluations of instructors which made me think I had potentiality as a contributor to my present field of study	21.9	24.7	51.8	+29.9**
Challenging and inspiring courses in my intended field of study	17.1	23.3	46.4	+29.3**
Discovering the high ability of students with whom I had to compete in my intended field of study	24.4	16.7	18.3	−6.1
Nothing I have experienced here has influenced my educational plans	11.0	38.4	4.2	−6.8*
A few instructors who alienated me	14.6	5.5	4.4	−10.2**
Negative evaluations of instructors	12.2	1.1	1.9	−10.3**
Dull and uninspiring courses in my intended field of study	29.3	12.9	18.3	−11.0*

[a] *N*s differ slightly from those shown in Table 5 because 101 students classified in other fields are included here and because of nonresponse to the question.

* $p < .05$.

** $p < .01$.

22% of those who did not change their aspirations. Experiences which reassured the student of his abilities in his field of study and "challenging and inspiring courses" were also mentioned much more frequently by those who raised their aspiration levels.

Student press. The results of scales designed to measure the norms and values of the student's classroom and living groups are not shown since none of the obtained differences for which predictions were made were statistically significant and only five of the nine were in the predicted direction. There was no evidence that student press influenced the aspirations of Merit Scholars and Certificate of Merit winners.

Factors Influencing Choice of Major Field

Identification with a stimulating teacher. Each student was asked to answer the question, "What college experiences have influenced your choice of career field?" by checking on a list of 12 alternatives those which applied. Half of the alternatives were designed to identify deterrents to remaining in an intended field of study, and half were designed to identify incentives attracting students into new fields of study. Among students recruited into one of the four major fields of study the two most popular "incentive responses" were: "evaluations of instructors which made me think I had potentiality as a contributor in my present field of study" (36%) and "a few instructors who inspired me to emulate them" (32%). The first was checked most frequently by students recruited into the biological sciences (44%) and the arts and humanities (40%). It was endorsed least frequently by students recruited into the natural sciences (29%). The second was endorsed most frequently (36%) by students trans-

ferring into the arts and humanities and least frequently (24%) by students transferring into the biological sciences.

Additional evidence that identification with a stimulating teacher is important was obtained by a more indirect question. Each student was asked to "list three faculty members (either within or outside your major department) in order of the degree to which you feel they contributed to your desire to learn in your chosen fields of study." The departmental affiliations of the faculty member listed in *first position* (as most influential) were then analyzed and related to changes in the student's major field of cencentration. Assuming that role models partly determine the student's choice of a major field, we expected that (*a*) students who maintain their interest in a field would more frequently name teachers in that field than would students who withdraw from the field, and (*b*) recruits into a field would more frequently name teachers in that field than would students who withdraw from the field. Both of these predictions were confirmed. Eighty-six percent of the stable students, and 73% of the recruits, named teachers in the field corresponding to their present major field of concentration, while only 13% of the students withdrawing from a field named teachers from that field. These data suggest that failure to find a stimulating teacher in one's initial field of concentration is a deterrent to the student's remaining in that field, and that he is frequently attracted into new fields by exposure to more stimulating teachers.

Deterrents to continued study in a field. The "deterrent responses" to the question "What college experiences have influenced your choice of career field?", mentioned above, suggest that many of the students who drop out of a field do so because their expectations of certain career-fields are proved to be incompatible with reality. Among students who withdrew from one of the four major fields, the most popular responses were: (*a*) "knowledge I obtained in college about my intended field of study" (53%), (*b*) "dull and uninspiring courses in my intended field of study" (37%), and (*c*) "information I obtained about opportunities to earn a livelihood in my intended field of study" (18%). The first two responses were most frequent among students withdrawing from the natural sciences (63 and 48% respectively) and least frequent among those withdrawing from the arts and humanities (44 and 18% respectively). Many talented students presumably chose their career fields without adequate occupational information. The results suggest that if attempts to arouse the interest of students in careers as scientists, engineers, or teachers do not provide accurate information about the duties, roles, and rewards to persons in these specialized fields, such efforts may boomerang.

Characteristics of the Stimulating Teacher

It will be recalled that each student was asked to list three faculty members (from any department) in order of the degree to which he felt they contributed to his desire to learn. The student was also asked to describe on a check list the behavior of the instructor who had the greatest influence upon his desire to learn. Selected characteristics of the course and instructor are listed in Table 7 in the order of frequency of mention. Students were nearly unanimous in reporting that influential teachers exhibited great enthusiasm when discussing their fields. Also, over 80% of the students reported that these teachers had taken a personal interest in them and their work.

Other traits attributed to these teachers varied significantly with the instructor's field of study; these results, like those reported in Table 5, suggest that teaching techniques effective in one field may not necessarily be effective in other fields. For example, influential teachers in the natural sciences, are distinguished from those in the humanities and social sciences by

TABLE 7

CHARACTERISTICS OF TEACHERS AND COURSES WHICH STUDENTS REPORT CONTRIBUTED MOST TO THEIR DESIRE TO LEARN

Characteristic of Course or Instructor	Percentage of Teachers Named in Each Field Who Exhibited Given Characteristic[a]				Percentage in All Four Major Fields (N = 1357)
	Natural Science Teachers (N = 593) (1)	Biological Science Teachers (N = 113) (2)	Social Science Teachers (N = 299) (3)	Arts and Humanities Teachers (N = 352) (4)	(5)
Exhibited great enthusiasm when discussing his field	96.0	96.4	95.8	96.8	96.2
Took a personal interest in me and my work	77.0	76.8	81.9	86.8	80.6
Frequently pointed to unsolved problems on which research is needed	75.3	87.4	85.0	64.7	75.7
Placed the highest priority upon helping the student to obtain a broad familiarity with knowledge and ideas in his field	79.1	89.2	72.5	62.6	74.3
Always discussed correct answers to tests	87.0	64.8	53.8	53.2	69.1
Course meticulously organized and progressing systematically from topic to topic	77.0	72.3	64.9	58.3	69.1
Was subject to change according to the *Ss*' needs and interests	58.7	61.3	62.6	71.3	63.1
Allowed time during class for extensive student discussion	51.2	54.1	67.6	78.0	62.0
Frequently encouraged qualified students to do graduate work	62.0	66.8	64.8	57.8	61.9
Provided personal evaluations which made me realize I might become a creative or productive worker in my field	49.9	54.1	61.5	76.4	59.7
Placed the highest priority upon developing the student's critical abilities	46.8	35.1	66.1	72.1	56.6
Required one or more term papers or laboratory projects	40.6	55.9	65.8	71.1	55.4
Usually outlined the day's lecture or discussion at the beginning of the class period	49.7	51.8	49.5	41.2	47.7
Often tried to relate the course to broad human values and goals	19.8	45.0	68.6	72.6	46.6
Meticulously documented sources of information	31.9	39.6	50.2	41.1	39.0
Encouraged competition among class members	45.7	29.1	31.6	30.0	37.1
Made special efforts to give recognition to students who did exceptional work	35.2	31.5	34.1	44.8	37.1
Was primarily a reading course	6.8	14.3	55.9	52.3	30.0
Was frequently critical of the ideas and work of other workers in his field	16.4	22.5	42.5	34.7	27.5

[a] In calculating percentages each teacher nomination was treated as if it designated a unique teacher even though occasionally students attending the same college nominated the same teacher. The *Ns* for percentages deviate slightly from those given above because some *Ss* failed to respond to particular checklist items; the average numbers of nonrespondents for Columns 1–4 were 20, 2, 8, and 11, respectively. The smallest differences found to be statistically significant ($p < .01$) between corresponding percentages in different columns were as follows: Columns 1-2, 12.1; Columns 1-3, 9.7; Columns 1-4, 9.2; Columns 2-3, 16.7; Columns 2-4, 14.0; Columns 3-4, 9.9.

their meticulous organization of courses, and by more frequent use of such pedagogical techniques as discussing correct answers to tests and encouraging competition among students. They also require term papers and laboratory projects less frequently than their colleagues in other fields. The influential teacher of biological sciences seems to be distinguished by stress upon giving the student a broad familiarity with knowledge and ideas in his field, and by less emphasis upon developing the student's critical abilities. At the same time, like the social science teacher, he more frequently points to unsolved problems on which research is needed.

Students describe stimulating teachers and courses in the arts, humanities, and social sciences in similar terms: more typical of these teachers than of teachers in other fields are "reading" courses, attempts to relate the course to broad human values and life goals, the use of classroom discussion and term papers or projects as pedagogic techniques, and stress upon developing the student's critical abilities. Social science teachers are distinguished by their relatively strong emphasis upon documentation of sources of information and their tendency to be critical of the ideas and work of colleagues. Teachers in the arts and humanities more frequently reassure students of their creative or productive potentialities, and make more systematic attempts to give recognition to students who do exceptional work. The traits which distinguish these teachers are consistent with the greater press for humanism and independence and the greater supportiveness of these faculties noted in Table 4.

Discussion

The greater success of the humanities and social sciences in recruiting new students may be attributed to several factors: the greater irreversibility of decisions *not* to major in the natural and biological sciences (Sibley, 1948); the image of teachers in the humanities and social sciences as being more enthusiastic, more tolerant of independence, and less pragmatic (Table 4); the greater use by the teacher of humanities and social sciences of pedagogic techniques which encourage students to play the role of a scholar or scientific investigator (Table 7).

Students majoring in the biological sciences more frequently ($p < .05$) raised their educational aspirations while in college (Table 3). Although the trait "frequently encouraged qualified students to do graduate work" was more often attributed to teachers in the biological sciences than to teachers in the arts and humanities (Table 7), the results do not suggest any simple explanation for this difference in "productivity."

The failure to replicate previous findings (Thistlethwaite, 1959a, 1959b) concerning the effects of student values and norms upon motivation to seek advanced training may be interpreted in terms of the personality characteristics of Merit finalists described by Holland (1959, 1960). Finalists have means on Gough's Ai scale (achievement via independence) which place them about two standard deviations above the norm for unselected students (Gough, 1957). Similarly, on Cattell's 16 PF scales (1957) finalists differ from the average college student in that they are more dominant (E), more radical (Q_1), and more self-sufficient (Q_2). Festinger's theory predicts that "if persons who are very divergent from one's own opinion or ability are perceived as different from oneself on attributes consistent with the divergence, the tendency to narrow the range of comparability becomes stronger" (1954, p. 133). A possible explanation is that many of these students have indeed narrowed the range of individual differences within which they compare themselves to their peers, and are consequently

iess susceptible to peer group norms than groups less highly select in terms of ability.

Departmental differences in faculty and student press have implications for evaluations of higher educational institutions (Knapp & Goodrich, 1952; Knapp & Greenbaum, 1953; Traxler, 1957; Thistlethwaite (1959a, 1959b). Table 4 suggests that departments differ considerably in what they consider to be the objectives of the undergraduate curriculum; differences in student norms and values suggest the output of a department may be a function of the characteristics of students it attracts. If colleges are to be compared on the basis of alumni achievements it is obviously necessary to adjust such assessments so that differences in the proportions of students in different fields of specialization are not confounded with differences in educational effectiveness.

Summary

This study reports the results of a follow-up, at the end of the third year of college, of 1500 talented students who had received recognition in a national scholarship competition. Student reports of faculty and student press were related to changes in the direction and level of career choices. The major results may be summarized as follows:

1. The natural sciences were relatively successful in retaining talented students who initially entered these fields, while the biological sciences had less holding power; both fields, however, attracted new students less frequently than the arts, humanities, and social sciences.

2. Experiences in college which encouraged the student's belief that he may be successful in a career-field, provided role models for imitation, or proved that his expectations about certain career-fields were inappropriate, were related to the stability and direction of change of study plans.

3. Faculties which students perceived as enthusiastic, warm and informal in their relationships with students (affiliative), and as stressing achievement, humanism, and independence were associated with more frequent changes in students' plans to seek advanced training in the arts, humanities, and social sciences.

4. Faculties which students perceive as enthusiastic and as not exercising great press for compliance were associated with more frequent changes in students' plans to seek advanced training in the natural and biological sciences.

5. Traits attributed to teachers who were most influential upon students varied considerably from field to field.

Implications for further research on the pressures and activities which characterize stimulating teachers, disciplines, and colleges are discussed.

REFERENCES

Cattell, R. B. *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign, Ill.: Inst. Pers. Ability Testing, 1957.

Festinger, L. A theory of social comparison processes. *Human Relat.* 1954, **7**, 117–40.

Gough, H. G. *Manual for the California Psychological Inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1957.

Gropper, G. L., & Fitzpatrick, R. *Who goes to graduate school?* Pittsburgh, Penn.: Amer. Inst. Res., 1959.

Holland, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *J. educ. Psychol.*, 1959, **50**, 135–42.

Holland, J. L. The achievement, aptitudes, and personalities of National Merit Scholars and Certificate of Merit winners. *Bull. Nat. Ass. Sec. Sch. Principals*, 1960, **44**, 100–104.

Iffert, R. E. Retention and withdrawal of college students. *U.S. Dept. Hlth., Educ., Welfare Bull.*, 1957, No. 1.

Knapp, R. H., & Goodrich, H. B. *Origins of American scientists.* Chicago, Ill.: Univer. Chicago Press, 1952.

KNAPP, R. H., & GREENBAUM, J. J. *The younger American scholar: His collegiate origins.* Chicago, Ill.: Univer. Chicago Press, 1953.

LEVINE, R. Estimating national norms for the SAT. Princeton, N.J.: Educational Testing Service, January 1958.

PACE, C. R., & STERN, G. G. An approach to the measurement of psychological characteristics of college environments. *J. educ. Psychol.*, 1958, **49**, 269–77.

SIBLEY, E. The recruitment, selection, and training of social scientists. *Soc. Sci. Res. Council Bull.*, 1948, No. 58.

STERN, G. G. *College characteristics index, preliminary manual.* Syracuse, N. Y.: Author, 1958.

THIELENS, W., JR. Some comparisons of entrants to medical and law school. In R. K. Merton et al. (Eds.), *The student-physician: Introductory studies in the sociology of medical education.* Cambridge, Mass.: Harvard Univer. Press, 1957.

THISTLETHWAITE, D. L. College environments and the development of talent. *Science*, 1959, **130**, 71–76. (a)

THISTLETHWAITE, D. L. College press and student achievement. *J. educ. Psychol.*, 1959, **50**, 183–91. (b)

TRAXLER, A. E. An appraisal of American Colleges on the basis of new graduates listed in *Who's Who in America.* Paper presented at the meeting of the AAAS, Indianapolis, 1957.

(Received January 27, 1960)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 235-239

FACULTY PERCEPTIONS OF STUDENT SUCCESS IN TEACHER EDUCATION

ROGER A. MYERS[1]
Ohio State University

Successful accomplishment in any complex social situation is difficult to define. When prediction is the goal, this problem of definition is especially crucial, as the OSS (1948) and Kelly and Fiske (1951) studies have demonstrated. In the area of teacher education, attempts to predict successful student accomplishment have been limited in scope and effectiveness by the lack of understanding of what constitutes performance that is judged successful. Investigators who needed acceptable indexes for evaluating their predictions have been forced to beg the question of definition and employ crudely unidimensional criteria, such as grade averages, which are assumed to represent the complicated concept of success. Although this is a widely accepted practice, most investigators realize that the object of their prediction is not the criterion measure itself, but rather the standard of performance which it represents.

One approach to the problem of defining successful accomplishment is suggested by the work of Pepinsky (1956). In his research on productive behavior he has conceptualized a social system which is made up primarily of two classes of people, actors and task-setters. While it is the actor whose behavior is studied and predicted, the role of the task-setter is a crucial one. It is he who decides what requirements are made of the actor, specifies the rules for meeting them, evaluates the actor's behavior, and rewards and punishes him accordingly. If one were to attempt to define successful accomplishment within such a social system, it would appear reasonable to begin by studying the task-setter in order to determine how he defines success. Since he is the one who determines success by his process of rewarding and punishing, the way in which he defines success is of paramount importance.

In the educational setting, studying the faculty member's views of successful student accomplishment is an infrequent occurrence. One recent exception is the work of Stern, Stein, and Bloom (1956). These investigators held a series of meetings with the faculty of a teacher education program during which they determined the things the faculty valued and the kinds of opinions and judgments they made about students. With this information the investigators constructed a hypothetical model of a student who would be regarded as successful. They then examined the students in the program, compared them to the model, and predicted how the faculty members would evaluate them. Although their predictions were extremely successful, the time and expense of their method for investigating faculty views of success limit is usefulness.

The present study was designed to investigate how the faculty of another teacher education program define successful student accomplishment. The specific purpose was to learn from the faculty members what characteristics of students they considered desirable and to construct dimensions that would represent the way these characteristics were organized in their perceptions. An attempt was made to conduct this investigation using research tools which are largely clerical and therefore more easily reproduced.

Method

Two-hour interviews were conducted with 12 members of a department of edu-

[1] Presently at the University of North Dakota.

cation faculty. Each of the faculty members interviewed was, in the opinion of the department chairman and other observers, a critical leader or representative of a focal position in the department. The interviews were unstructured except for the interviewer's request that each faculty member describe, in his own words, the behavior of a good student and a poor student in the department. The sample size of 12 was determined when, in the judgment of the interviewer, redundancy occurred in the interview and the addition of faculty members to the sample produced no new descriptive material.

From typescripts of the interviews the investigator extracted 280 statements which the faculty members had used in describing the good and poor students. Of these 280 statements, 169 were finally judged to be usable on the basis of clarity, objectivity, and general applicability. The descriptive statements were then given to eight judges who sorted them into 15 meaningful categories. These categories were given the following titles: hostility, flexibility, intellectual curiosity, general intelligence, substantive preparation, autonomy, rejection, order and detail, nurturance, self-confidence, energy and enthusiasm, growth potential, sociability, dominance, and dependence.

Items from each category were systematically spaced throughout a checklist so that statements from a single category did not confront the reader more frequently than once in ten. The order of the pages of the checklist was randomized in an attempt to control error which might have resulted from the position of a statement on the checklist.

Of the 61 members of the department faculty who were on duty, 6 had no contact with the undergraduate students and felt that they could not appropriately respond to the check list. Each of the remaining 55 members was given two copies of the check list with instructions to describe one good student and one poor student whom he knew. The students were to be in their third or fourth year of college and in the teacher education program. The instructions for choosing a good student suggested, "... one who closely approximates the realization of the objectives you hold in your teaching and work with students." Instructions for choosing a poor student suggested, "... one who has managed to remain in college but who has not developed in accordance with your goals for students." The descriptions were accomplished by assigning a rating on a five-point scale to each of the descriptive statements, according to the degree to which it was characteristic of the student being described. Fifty of the 55 faculty members returned completed checklists.

The resulting ratings were factor analyzed by the Wherry-Winer method for use with large numbers of items (1953). First, category scores were obtained by summing the ratings of the descriptive statements in each category. Then an item analysis was performed for each category. As a result of these item analyses, two of the categories, dominance and dependence, were dropped and the items in them were assigned to an unclassified pool. The remaining 13 categories, used as subtests, were intercorrelated. The resulting matrix included some correlations that were unusually high and suggested that some of the subtests did not meet the requirement for linear independence imposed by the method. A test for linear independence resulted in the dropping of five more categories and the assignment of more items to the unclassified pool. These categories were flexibility, general intelligence, substantive preparation, energy and enthusiasm, and growth potential.

The remaining matrix included eight categories: hostility, intellectual curiosity, autonomy, rejection, order and detail, nurturance, self-confidence, and sociability. A multiple-group centroid analysis of this matrix yielded a general factor and three group factors. From these factors a trans-

formation matrix was derived which, when multiplied by previously computed correlations between statements and categories, produced orthogonal loadings for each of the statements on each of the factors.

Results

The four factors obtained were assumed to represent dimensions of successful student behavior in the perception of the faculty members. Each factor was identified by the descriptive statements which had loadings on it. Because of the limitation of space it is impractical to list all of the statements and factor loadings here. Table 1 includes a sample of these chosen from the statements with the highest factor loadings.[2]

Factor I is a general factor on which most of the descriptive statements have high loadings. It was hypothesized that this factor represents a general perceptual orientation of the faculty toward acceptable student behavior. It appears to be a generally shared stereotype of students which arises from actual and perceived similarities among the statements. The strength of the factor suggests a high degree of commonness among the faculty members' views of student success. In a similar study, Stoltz (1959) has pointed out the contribution of halo effect to this kind of factor.

The name given to this factor was General Acceptability. As a dimension of student behavior it is assumed to represent some kind of global perception on the part of the faculty members.

Factor II was named Readiness to Learn. It contains statements which describe behavior that the judges categorized as intellectual curiosity, autonomy, self-confidence, energy, general intelligence, and substantive preparation. It is characterized by the student who comes to the learning situation with all of his credentials in order. He has made a habit of doing what is expected of him and in addition, when left to his own devices, he will use his time and energy profitably. He is not only capable of dealing with higher level abstractions, but also enjoys going beyond the immediate in his search for intellectual stimulation. Finally, he is free from neurotic fears which might decrease the vigor of his search for knowledge and complicate his relationships with his professors.

It is apparent that this dimension is highly related to the general capacity variable that is widely recognized as a major determinant of success in any academic program. Added to this general capacity, however, are strong motivation relevant to the goals of the department and personal security which permits unencumbered professor-student relationships.

Since Factor III is composed largely of statements that the judges originally classified as nurturance and sociability, it was given the name of Social Sensitivity. The behavior described occurs most frequently in the student teaching situation and in the student's daily relationships with his peers. It includes such things as initiating and directing group action that is facilitative to learning; observing, understanding, and helping pupils and peers; being tolerant of the weaknesses of others; and so on. The emergence of this dimension coincides to some extent with the findings of Ohlsen and Schultz (1954), Tyler (1954), and Stern, Stein, and Bloom (1956).

Factor IV is composed of negatively loaded statements from the hostility and rejection categories of the judges and positively loaded statements from the category called order and detail. The negatively loaded statements include behavior that suggests the acting out of rebellious feelings toward the authority figures represented by the professors, lack of respect for the institution and its goals, and a

[2] A complete list of the statements and their factor loadings is included in R. A. Myers, "A factor analytic study of faculty views of student success." Unpublished doctoral dissertation, Ohio State University, 1959, pp. 46–51, 56–58, and 62–64.

TABLE 1

FACTOR LOADINGS FOR A SAMPLE OF CHECK LIST STATEMENTS

Statement	Factor			
	I	II	III	IV
Is confident in examinations	65	66	−06	08
Will attend a university sponsored lecture even if it is not required	62	65	−02	06
Has a well-developed vocabulary	70	63	05	05
Writes papers with coherent conclusions	72	62	06	16
Can illustrate his abstractions with ease	69	60	04	04
Will question broad generalities, even outside his own area of knowledge	63	59	00	−04
Can be critical of a professor's point of view	47	59	01	−20
Has arrived at his own value system	53	58	02	13
Has good grammatical habits	70	56	06	09
Is eager to have his work evaluated	68	56	−06	25
Has a broad general education	65	55	03	07
Has the ability to relate to children	66	−18	53	11
Gets respect from his pupils or peers	72	04	45	29
Has patience with slow learners	58	−25	44	13
Can get to the experience level of students he is teaching	70	−11	43	12
Is alert to facial expressions of pupils or peers	62	−06	43	−06
Says things in class that get the rest of the group involved	68	16	42	−08
Can be firm and decisive with pupils or peers	69	07	41	−04
Has a great deal of warmth in interpersonal relationships	65	00	40	30
Is able to get a group interested in a project	75	03	38	12
Projects himself into the thinking of his pupils or peers	68	19	32	17
"Pops off" in class	−22	08	02	−70
Wants to do things his way, regardless of the rules	−38	08	−05	−70
Is generally "fed up" with the College of Education	−47	00	−01	−66
Obstructs class progress by thinking up reasons why things won't work	−37	02	01	−63
Thinks he knows more than his professors	−36	15	−03	−62
Expresses negative feelings about course material	−23	10	06	−60
Is punctual	46	38	11	58
Complains about classes being too large	−32	02	16	−57
Is hostile to his fellow students	−53	−18	−01	−52
Is uncooperative in group projects	−41	05	−18	−48
Turns in required work on time	56	38	18	46
Shows great loyalty and deference to authorities	−14	−07	−09	44

Note.—All decimals are omitted.

characteristically hostile and uncooperative orientation to others. The positively loaded statements describe behavior that suggests compliance with daily requirements, orderly work habits, and attention to details. The combination of these two kinds of statements seems to define a dimension that has to do with controlling aggressive emotions vs. expressing them. The factor was given the name of Impulse Control.

SUMMARY

Fifty members of a department of education faculty described the behavior of good and poor students in a teacher education program using a check list of 169 de-

scriptive statements. A Wherry-Winer factor analysis of these statements resulted in the extraction of four factors which were assumed to represent dimensions of student behavior as perceived by faculty members. The factors have been titled General Acceptability, Readiness to Learn, Social Sensitivity, and Impulse Control.

These dimensions help to define the standard of performance that the faculty of one teacher education program expect of their students. Further study could lead to the development of ways of evaluating students on these dimensions for the purpose of arriving at a multiple criterion of success.

REFERENCES

KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. Michigan Press, 1951.

OHLSEN, M. M., & SCHULTZ, R. E. A study of variables for use in selection. *J. teach. Educ.*, 1954, **5**, 279–282.

OSS ASSESSMENT STAFF. *Assessment of men.* New York: Rinehardt, 1948.

PEPINSKY, H. B. Productivity in the university. *Personnel guid. J.* 1956, **35**, 134–139.

STERN, G. G., STEIN, M. I., & BLOOM, B. S. *Methods in personality assessment.* Glencoe, Ill.: Free Press, 1956.

STOLTZ, R. E. Factors in supervisors' perceptions of physical science research personnel. *J. appl. Psychol.*, 1959, **43**, 256–258.

TYLER, F. T. The prediction of student teaching success from personality inventories. *Univer. Calif. Publ. Educ.*, 1954, **11**, 233–314.

WHERRY, R. J., & WINER, B. J. A method for factoring large numbers of items. *Psychometrika*, 1953, **18**, 161–179.

(Received February 11, 1960)

Journal of Educational Psychology
1960, Vol. 51, No. 4, 240-244

CHANGES IN SCORES ON THE NORTHWESTERN MISCONCEPTIONS TEST IN SIX ELEMENTARY PSYCHOLOGY COURSES

W. J. McKEACHIE
University of Michigan

In 1950 the Division on the Teaching of Psychology formed a Committtee on Research in the Teaching of Psychology[1] which in its three years of existence sponsored a number of APA programs, surveyed the need for publication space for research in the teaching of psychology, and stimulated the writing of an article surveying research on the teaching of psychology (Birney & McKeachie, 1955). The committee also discussed the need for more comprehensive achievement tests for which normative data might be collected. While no agreement on construction of a test was reached, the members of the committee did decide to administer the Northwestern Misconceptions Test (Holley & Buxton, 1950) to their classes to secure normative data for it. The test consists of 100 true-false items collected by Buxton from numerous sources, selected on the basis of item-analysis and difficulty. There are an equal number of true and false correct responses. Newman, Duncan, Bell, and Brandt (1952) report a split-half reliability of .71. The test is apparently frequently used in general psychology courses, at least as part of the introduction to the course.

By giving the test at a number of institutions we hoped not only to gather normative data on the types and number of misconceptions college students hold, but also to determine the effect of different introductory courses upon changing misconceptions.

[1] Members of the committee were F. K. Berrien, Everett Bovard, C. Ray Carpenter, Carl P. Duncan, Paul Fields, B. Von Haller Gilmer, Wilbert S. Ray, Clarence Young, and W. J. McKeachie, Chairman.

Test Administration

The test was administered to the elementary psychology classes at Carnegie Institute of Technology, Colgate, Michigan, Northwestern, Pennsylvania State, and Trinity, half the sections taking the test at the beginning of the semester and all sections taking it at the end of the semester.

Results

Pre-Post course gains. Since there were no reliable differences between end-test results for those who had had the pretest and those who had not, the total results will be reported for students who took both pre- and posttests at each institution. (See Table 1.)

Although the overall gains at the participating institutions may be of interest to other institutions using the test, most psychologists will be more interested in responses to individual items, since these give some clues to the beliefs our students have and the effect introductory courses are likely to have. Such results would be most convincing if the introductory courses represented here had similar effects.

Table 2 shows pre-posttest scores for selected items. Unfortunately, no item showed significant gains at all schools, nor did any item show significant decrements from pre- to posttest at all schools. Item 85 did show a loss in responses at three institutions, and only one school recorded a pre-posttest gain. Item 39 showed significant losses at two institutions. Item 25 showed a significant gain at one school and a significant loss at another. Only Items 22 and 65 showed significant pre- to posttest gains

at more than one school. When items were grouped by content area, such as perception, motivation, etc., there were no striking differences in correct responses between categories.

Some items in Table 2 are identified as representing the four factors isolated by Holley and Buxton in an inverted factor analysis (Q technique) of 20 students at Northwestern. They ran biserial correlations for passing an item and individual factor scores, thus obtaining the items which most clearly differentiated individuals high and low on a particular factor. These items were then used to identify the factor. It is clear from Table 2 that average gains for items in Factor I and those with positive loading on Factor II are negligible. This might be accounted for by a preponderance of easy items, since over 70% of the students passed these items on the pretest. However, this is not a sufficient explanation since 77% of the responses to items in Factor III were correct on the pretest and significant gains still occurred.

TABLE 1

MEAN SCORES ON MISCONCEPTIONS TEST

Institution	Pretest	Posttest	Gains[a]	N
Carnegie Tech.	69.2	75.6	6.4	96
Colgate	60.4	64.5	4.1	40
Michigan	68.8	73.7	4.9	57
Northwestern[b]	61.7	72.0	10.3	154
Pennsylvania State	61.1	64.8	3.7	113
Trinity	64.0	70.4	6.4	53

[a] All gains are statistically significant at the .001 level.

[b] At Northwestern correlation between pre- and posttest was .35.

DISCUSSION

If one of the goals of the elementary psychology course is to eliminate misconceptions about psychology and the Northwestern test is valid, instructors in none of the courses represented in our study can be very proud of their accomplishment. The typical student who has completed one of these courses retains at the end of the course about 30 of the 100 misconceptions, or about 85% of those with which he began his course in psychology. Even at the end of the course students believe that:

Item 27. The sense organs of touch, in a person with normal vision, are just as sensitive as those in a blind person, is false.

Item 53. Rats, cats, and dogs have the power to reason, is false.

Item 72. After graduation from college, the average Phi Bete makes more money than the average non-Phi Bete, is false.

Item 74. College men are less promiscuous sexually than men of equal age but lesser education, is false.

Item 91. Psychiatrists are defined as medical men who use psychoanalysis, is true.

Validity. Different psychologists will view the importance of these beliefs differently. In fact, it is quite possible that psychologists themselves would not reach unanimity on the correct answers to the items. To compare student responses with those of instructors, the staff of the general psychology courses at the University of Michigan was asked to rate each item on the following scale:

a. I disagree with the key.
b. I don't care how my students answer.
c. I'd be mildly unhappy if my students missed this.
d. I'd be quite unhappy if my students missed this.

The only item which was scored "d" by all 11 respondents was Item 26. Only seven other items were rated "c" or "d" by all of the 11 instructors. A majority of the instructors chose "a" for one item. Of the 24 items which were passed by less than 60% of the students at the end of the course, one was the item with the questionable key and the rest were items which at least 40% of the instructors rated "b". Thus, there is apparently little cause for concern about these misconceptions. On the other hand, even though most students come into gen-

eral psychology with correct beliefs about the 8 items regarded as most important by the instructors, we have relatively little effect on those students who do hold these misconceptions. On these 8 items, 80.5% of the student responses were correct on the pretest and 82% were correct on the posttest.

What does this add up to? The author's first reaction is one of disillusionment—not only with the extent to which our teaching generalizes to everyday beliefs, but also

TABLE 2

CHANGES IN RESPONSES TO SELECTED ITEMS ALL INSTITUTIONS COMBINED (N = 513)

Item	Content	Scored by Original Northwestern Key	Pretest % Correct	Posttest % Correct
A. Items reported by Holley and Buxton as having high loadings on their first factor				
53	Rats, cats, and dogs have the power to reason.	*T*	40	34
87	The printing on this page is upside-down on your retina.	*T*	93	93
91*	Psychiatrists are defined as medical men who use psycho-analysis.	*F*	41	43
97	There is little that psychology can do for the *normal* person.	*F*	93	92
B. Items reported by Holley and Buxton as having high loadings on their second factor				
2	Cats can see in complete darkness.	*F*	86	92
47	A dog can sense impending disaster better than a man.	*F*	47	60
71	Half the people in this country are below average in intelligence.	*T*	54	57
76	The unconscious mind is located just above the roof of the mouth, directly back of the nose.	*F*	83	85
C. Items reported by Holley and Buxton as having high loadings on their third factor				
6*	A child is born with a sense of good and evil, this is his conscience.	*F*	85	91
7*	The majority of adult criminals are feeble-minded or very nearly so.	*F*	83	86
37	A person who won't look you in the eye is probably untrustworthy.	*F*	89	93
63	Being spanked may be pleasurable to a child.	*T*	58	78
D. Items reported by Holley and Buxton as having high loadings on their fourth factor				
27	The sense organs of touch, in a person with normal vision, are just as sensitive as those in a blind person.	*T*	32	33
46	It is probable that man's instinct to fight is the fundamental cause of wars.	*F*	76	87
65	The average infant would learn to walk two months earlier than he does, if he were given proper training.	*F*	51	69
77	Children memorize much more easily than adults.	*F*	37	44

TABLE 2.—*Continued*

Item	Content	Scored by Original Northwestern Key	Pretest % Correct	Posttest % Correct
E. Other items referred to in text				
11*	The most important consideration to the industrial worker is the amount of money a job pays.	*F*	70	67
17*	Most college students are not susceptible to propaganda.	*F*	87	90
22	Very bright children tend to be at least equal to children of average intelligence, with respect to physical development.	*T*	64	82
25	The very bright child tends to be at least equal to children of average intelligence, with respect to emotional adjustment.	*T*	56	70
26*	All men are created equal in capacity for achievement.	*F*	83	88
39	The I.Q. can be changed, to a significant degree.	*T*	52	49
62*	Worry can cause a truly physical disease.	*T*	87	90
67*	There is no evidence that Negroes are inherently (naturally) less intelligent than whites.	*T*	86	91
72	After graduation from college, the average Phi Bete makes more money than the average non-Phi-Bete.	*T*	28	32
74	College men are less promiscuous sexually than men of equal age but lesser education.	*T*	41	37
85	A properly-trained psychologist (not a quack) can tell something about the personality of an individual from his handwriting alone.	*T*	64	42

Note.—The data from Northwestern are not included in this table. Numbers passing pre- and posttest at the other institutions were simply added to derive this table ($N = 359$).

* These items were rated as important by the Michigan staff.

with the validity of the Northwestern test as a measure of these beliefs.[2] However, even though one can rationalize the relatively small gains by attacking the significance, wording, or item form of the Northwestern test, this is not a very satisfying rationalization; for other studies using other items have arrived at similarily depressing conclusions about application and generalization of learning in psychology courses (Horrocks & Nagy, 1948; Magaret & Sherriffs, 1947). Clearly if we're really serious about aiming for generalization of or application of psychological performance, we need to investigate more carefully what types of teaching or content are most effective in producing gains in this area.

[2] Although the Michigan staff has some doubts about the value of a misconceptions test of this type, Byron Groesbeck and the author have developed a 20-item type test consisting of 14 items from this test and six new items. Thirteen of these meet the following criteria: (*a*) content considered worth knowing by at least 70% of the Michigan staff, (*b*) passed by less than 70% of the students on a pretest, (*c*) item validity (against total score) of .20 or higher. The other seven were those rated most important by the Michigan staff. Mimeographed copies of the test are available from the author.

Summary

The Northwestern Misconceptions Test was administered at the beginning and end of the elementary psychology course at six institutions.

Although all of the institutions produced

gains in total test score, gains or losses on individual items varied from institution to institution.

Many of the items upon which students showed little improvement were regarded by faculty members as unimportant. However, gains on items rated as important were also very small.

REFERENCES

BIRNEY, R., & MCKEACHIE, W. J. The teaching of psychology: A survey of research since 1942. *Psychol. Bull.*, 1955, **51**, 51-60.

DUNCAN, C. P., O'BRIEN, R. B., MURRAY, D. C., DAVIS, L., & GILLILAND, A. R. Some information about a test of psychological misconceptions. *J. gen. Psychol.*, 1957, **56**, 257-260.

HOLLEY, J., & BUXTON, C. A factorial study of beliefs. *Educ. psychol. Measmt.*, 1950, **10**, 400-410.

HORROCKS, J. E., & NAGY, G. The relationship between the ability to make a diagnosis and to select appropriate remedial procedures. *J. gen. Psychol.*, 1948, **38**, 139-145.

MAGARET, ANN, & SHERRIFFS, A. C. Personality factors influencing learning in the first course in psychology. *J. gen. Psychol.*, 1947, **37**, 67-71.

NEWMAN, S. E., DUNCAN, C. P., BELL, G. B., & BRADT, K. H. Predicting student performance in the first course in psychology. *J. educ. Psychol.*, 1952, **43**, 243-247.

(Received February 17, 1960)

SCHOOL PSYCHOLOGISTS AT MID-CENTURY

Edited by

NORMA E. CUTTS

The Report of the Thayer Conference, published in 1955, covered an intensive study of the development of school psychology. Because of continuing demand, the Report has been reprinted.

230 pages. Price $2.75

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington 6, D. C.

GRADUATE EDUCATION IN PSYCHOLOGY

Report of the Conference on Graduate Education in Psychology, sponsored by the Education and Training Board of the American Psychological Association and supported by a grant from the National Institute of Mental Health, U. S. Public Health Service; held at Miami Beach, Florida, November 29 to December 7, 1958.

Prepared by the Editorial Committee:

Anne Roe, *Chairman*,

and

John W. Gustad, Bruce V. Moore,

Sherman Ross, and Marie Skodak

Price $1.50

THE AMERICAN
PSYCHOLOGICAL ASSOCIATION
Dept. Grad
1333 Sixteenth Street, N.W.
Washington 6, D. C.

SENT TO YOU UPON REQUEST

New

1950–1961 Catalogs

Of Special Interest to Psychologists

WESTERN PSYCHOLOGICAL SERVICES

Box 757, Dept. EP

Beverly Hills,

California